



## **REMARKS**

Claims 1-4, 13-17, 19-20, and 30-33 are currently under consideration. Claims 5-12, 18, and 21-29 were previously canceled. No amendments have been made to the claims in this response. Applicants address below each issue raised by the Office in the Office Action dated June 16, 2005.

### **Incorporation by Reference**

On page 3 of the Office Action, the Office requests Applicants to provide a list of the applications from the list of patent applications incorporated by reference on pages 513-533 that are related to the elected subject matter. In response, Applicants respectfully submit that the following applications relate to HEMCM42:

- U.S. Provisional Application No. 60/048,190, filed May 30, 1997 (listed on page 516, line 33 of the instant specification);
- U.S. Application No. 09/189,144, filed Nov. 10, 1998 (listed on page 531, line 23 of the instant specification), which is now abandoned and which claims priority to U.S. Provisional Application No. 60/048,190 above; and
- PCT/US98/10868 (published as WO 98/54206), filed May 28, 1998 (listed on page 531, line 24 of the instant specification), which also claims priority to U.S. Provisional Application No. 60/048,190.

### **35 U.S.C. § 112, Second Paragraph**

On page 4 of the Office Action, claims 1-4, 13-17, 19, 20, and 30-33 are rejected under 35 U.S.C. § 112, second paragraph, because the Office alleges that the “definition of ‘protein encoded by HEMCM42’ is vague and unclear as ‘HEMCM42’ is not

defined either in the claims or in [the] specification. There is no sequence associated with this abbreviation." Applicants respectfully traverse.

In the table on page 59, row 23 of the instant specification, the amino acid sequence of HEMCM42 is disclosed as corresponding to Genseq Accession No. W73409 (copy enclosed) and originating from WO 98/54206 (copy enclosed). This amino acid sequence was first entered into Genseq February 19, 1999, before the filing date of the instant application. Moreover, on page 30, WO 98/54206 discloses HEMCM42 as Gene No. 13 and its nucleotide and amino acid sequences as SEQ ID NOs: 23 and 60, respectively, in the sequence listing. WO 98/54206 was published on December 3, 1998, also before the filing date of the instant application. Therefore, the nucleotide and amino acid sequences of HEMCM42 were publicly available and known to one skilled in the art prior to the filing date of the instant application. Moreover, the instant application on page 59 provides clear guidance to one skilled in the art where to look for those sequences.

For the reasons above, Applicants deem it unnecessary to incorporate the HEMCM42 sequences into the specification as they were well known and understood by those skilled in the art at the time this application was filed. A patent need not teach, and preferably omits, what is well known in the art. *Hybritech Inc. v. Monoclonal Antibodies, Inc.*, 231 U.S.P.Q. 81, 94 (Fed. Cir. 1986). Nonetheless, in the interest of furthering prosecution, Applicants have amended the instant specification to add the nucleotide and amino acid sequences of HEMCM42, which were disclosed as SEQ ID NOs: 23 and 60, respectively, in WO 98/54206. The amendments do not constitute new matter because the HEMCM42 sequences were incorporated by reference to WO

98/54206, as indicated on page 11, lines 9-18, and on page 59, row 23 of the specification, and the amendatory material consists of the same material incorporated by reference in the present application. See Declaration of Charles E. Van Horn. The nucleotide and amino acid sequences of HEMCM42 have been added in the substitute sequence listing and in the specification on page 59 as SEQ ID NOs: 2278 and 2279, respectively.

As amended, the specification now includes the nucleotide and amino acid sequences of HEMCM42 as SEQ ID NOs: 2278 and 2279, respectively. Therefore, the term "HEMCM42" is not vague or unclear. Applicants respectfully request that the rejection be withdrawn.

**35 U.S.C. § 101/112, First Paragraph - Utility**

On page 5 of the Office Action, claims 1-4, 13-17, 19, 20, and 30-33 are rejected under 35 U.S.C. § 101 because "[t]here is no information regarding the nature of nucleic acid abbreviated as 'HEMCM42', nor there is any indication regarding utility of the particular 'protein encoded by HEMCM42'." The claims are also rejected under 35 U.S.C. § 112, first paragraph, because the claimed invention is allegedly "not supported by either a credible asserted utility or a well established utility." Applicants respectfully disagree.

The instant specification, under column "Preferred Indication Y" in Table 1 on page 59, discloses "cancer" as a specific and credible utility for HEMCM42. The instant specification further discloses that "recitation of 'Cancer' in the 'Preferred Indication Y' column indicates that [t]he corresponding Therapeutic protein, fusion protein containing the Therapeutic protein, and fragments and variants thereof, may be used for example,

to diagnose, treat, prevent and/or ameliorate diseases and/or disorders relating to neoplastic diseases . . . “ (page 11, line 30 to page 12, line 2). The neoplasm may be located in, for example, the “colon, abdomen, bone, breast, digestive system, liver, pancreas, prostate, peritoneum . . .” (page 12, lines 3-11). The fusion protein may also be used to “diagnose, treat, prevent, and/or ameliorate a pre-neoplastic condition” such as hyperplasia, metaplasia, and/or dysplasia (page 12, lines 13-20), or benign dysproliferative disorders such as benign tumors, fibrocystic conditions, and/or tissue hypertrophy (page 12, lines 21-27).

In addition, HEMCM42 protein has greater than 80% identity to the TWEAK receptor over the entire length of the protein and 100% identity over amino acids 1 to 106 of the TWEAK receptor (see Exhibit A attached). Thus, HEMCM42 protein has an amino acid sequence that is significantly related to the TWEAK receptor. The TWEAK receptor, also known as Fn14, was a well-characterized protein in the art and was also known to be useful for diagnosing, treating, preventing, and/or ameliorating cancer. See Feng et al., “The FN14 Immediate-Early Response Gene is Induced During Liver Regeneration and Highly Expressed in Both Human and Murine Hepatocellular Carcinomas,” *Am. J. Pathol.* **156**(4):1253-1261 (April 2000) (“the Feng reference”).

The Feng reference discloses that the Fn14 gene was overexpressed in poorly differentiated, human hepatocellular carcinoma (HCC) cell lines compared to cell lines derived from normal liver tissue. See, sentence bridging pages 1258 and 1259. In addition, the authors found that Fn14 expression was elevated in mice coexpressing *c-myc* and TGF- $\alpha$ , which serve as a model of hepatocyte growth and neoplasia. See, paragraph bridging pages 1256 and 1257; and page 1260, left column, first full

paragraph. These observations led the authors to conclude that “Fn14 gene activation may be associated with liver tumorigenesis . . .” (page 1259, left column) and that “a high level of Fn14 expression in HCC may promote cell detachment from the primary tumor, thus contributing to intra- or extrahepatic metastasis” (page 1260, sentence bridging left and right columns).

The use of the TWEAK receptor in inhibiting tumor growth was confirmed by Wiley (WO 01/45730) (copy enclosed). See, e.g., p15, lines 28-30 (“treatment is advantageously administered in order to prevent the onset or the recurrence of a disease or condition mediated by angiogenesis. . . disease and conditions mediated by angiogenesis include but are not limited to. . . malignant and metastatic conditions . . .”); page 16, lines 1-2 (“[t]he present invention can also be used to treat malignant and metastatic conditions such as solid tumors . . .”); page 16, lines 5-6 (“[o]ther diseases and conditions that can be treated according to the present invention include benign tumors and preneoplastic conditions . . .”).

Taken together, the specification and the knowledge in the art at the time of filing of the instant application provide a credible and well-established utility for the HEMCM42 protein. Therefore, withdrawal of the rejection is requested.

**35 U.S.C. § 112, First Paragraph - Enablement**

On pages 5-6 of the Office Action, claims 1-4, 13-17, 19, 20, and 30-33 are rejected under 35 U.S.C. § 112, first paragraph, as allegedly failing to comply with the enablement requirement. The Office contends that because “[t]here is no sequence associated with this abbreviation [HEMCM42] . . . the claims contains [sic] subject matter which was not described in the specification in such a way as to enable one

skilled in the art to which it pertains, or with which it is most nearly connected, to make the invention.” Applicants respectfully traverse.

As discussed above under the section “35 U.S.C. § 112, Second Paragraph,” the specification, as filed, discloses in the table on page 59, row 23, that the amino acid sequence of HEMCM42 corresponds to Genseq accession no. W73409 and originates from WO 98/54206. WO 98/54206 discloses the nucleotide and amino acid sequences of HEMCM42 as SEQ ID NOs: 23 and 60, respectively. WO 98/54206 was published December 3, 1998, and Genseq accession no. W73409 was first entered February 13, 1999. Therefore, both the nucleotide and amino acid sequences were publicly available prior to the filing date of the instant application to enable one skilled in the art to make the invention. Moreover, the instant specification has been amended to incorporate the nucleotide and amino acid sequences of HEMCM42 as SEQ ID NOs. 2278 and 2279, respectively. Therefore, withdrawal of the rejection is requested.

Applicants respectfully request that this Amendment under 37 C.F.R. § 1.116 be entered by the Examiner, placing claims 1-4, 13-17, 19-20, and 30-33 in condition for allowance. Applicants submit that the proposed amendments to the specification do not raise new issues or contain any new matter.

In view of the foregoing remarks, Applicants request the entry of this Amendment, the Examiner's reconsideration and reexamination of the application, and the timely allowance of the pending claims.

Please grant any extensions of time required to enter this response and charge any additional required fees to our deposit account 06-0916.



Respectfully submitted,

FINNEGAN, HENDERSON, FARABOW,  
GARRETT & DUNNER, L.L.P.

Dated: August 26, 2005

By: Charles E Van Horn  
Charles E. Van Horn  
Reg. No. 40,266

**Attachments:**

- Declaration of Charles E. Van Horn
- Substitute Sequence Listing (paper and computer readable copy)
- Statement supporting submission of Substitute Sequence Listing
- Genseq Accession No. W73409
- WO 98/54206
- Exhibit A
- Feng et al., *Am. J. Pathol.* **156**(4):1253-1261 (April 2000)
- WO 01/45730



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

|   |  |   |   |
|---|--|---|---|
| (51) International Patent Classification <sup>6</sup> :<br><b>C07K 1/00, C07H 21/04</b>   |  | <b>A1</b>   | (11) International Publication Number: <b>WO 98/54206</b>       |
|   |  |   | (43) International Publication Date: 3 December 1998 (03.12.98) |
| (21) International Application Number: <b>PCT/US98/10868</b>  |  | Court, Carlsbad, CA 92009 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US).   |   |
| (22) International Filing Date: <b>28 May 1998 (28.05.98)</b>   |  |   |   |
| (30) Priority Data:   |  |   |   |
| 60/044,039 30 May 1997 (30.05.97) US<br>60/048,093 30 May 1997 (30.05.97) US<br>60/048,190 30 May 1997 (30.05.97) US<br>60/050,935 30 May 1997 (30.05.97) US<br>60/048,101 30 May 1997 (30.05.97) US<br>60/048,356 30 May 1997 (30.05.97) US<br>60/056,250 29 August 1997 (29.08.97) US<br>60/056,296 29 August 1997 (29.08.97) US<br>60/056,293 29 August 1997 (29.08.97) US   |  |   |   |
| (71) Applicant (for all designated States except US): <b>HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US).</b>   |  |   |   |
| (72) Inventors; and   |  | (74) Agents: <b>BROOKES, A., Anders et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 20850 (US).</b>  |   |
| (75) Inventors/Applicants (for US only): <b>RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hill Road, Laytonsville, MD 20882 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). DILLON, Patrick, J. [US/US]; 1055 Snipe</b>   |  |   |   |
|   |  | (81) Designated States: <b>AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</b> |   |
|   |  | <b>Published</b><br><i>With international search report.</i><br><i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>   |   |
| (54) Title: <b>32 HUMAN SECRETED PROTEINS</b>   |  |   |   |
| (57) Abstract   |  |   |   |
| <p>The present invention relates to 32 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.</p> |  |   |   |

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

|    |                          |    |  |    |  |    |                          |
|----|--------------------------|----|--|----|--|----|--------------------------|
| AL | Albania                  | ES | Spain                                    | LS | Lesotho                                      | SI | Slovenia                 |
| AM | Armenia                  | FI | Finland                                  | LT | Lithuania                                    | SK | Slovakia                 |
| AT | Austria                  | FR | France                                   | LU | Luxembourg                                   | SN | Senegal                  |
| AU | Australia                | GA | Gabon                                    | LV | Latvia                                       | SZ | Swaziland                |
| AZ | Azerbaijan               | GB | United Kingdom                           | MC | Monaco                                       | TD | Chad                     |
| BA | Bosnia and Herzegovina   | GE | Georgia                                  | MD | Republic of Moldova                          | TG | Togo                     |
| BB | Barbados                 | GH | Ghana                                    | MG | Madagascar                                   | TJ | Tajikistan               |
| BE | Belgium                  | GN | Guinea                                   | MK | The former Yugoslav<br>Republic of Macedonia | TM | Turkmenistan             |
| BF | Burkina Faso             | GR | Greece                                   |    |  | TR | Turkey                   |
| BG | Bulgaria                 | HU | Hungary                                  | ML | Mali   | TT | Trinidad and Tobago      |
| BJ | Benin                    | IE | Ireland                                  | MN | Mongolia                                     | UA | Ukraine                  |
| BR | Brazil                   | IL | Israel                                   | MR | Mauritania                                   | UG | Uganda                   |
| BY | Belarus                  | IS | Iceland                                  | MW | Malawi                                       | US | United States of America |
| CA | Canada                   | IT | Italy                                    | MX | Mexico                                       | UZ | Uzbekistan               |
| CF | Central African Republic | JP | Japan                                    | NE | Niger  | VN | Viet Nam                 |
| CG | Congo                    | KE | Kenya                                    | NL | Netherlands                                  | YU | Yugoslavia               |
| CH | Switzerland              | KG | Kyrgyzstan                               | NO | Norway                                       | ZW | Zimbabwe                 |
| CI | Côte d'Ivoire            | KP | Democratic People's<br>Republic of Korea | NZ | New Zealand                                  |    |                          |
| CM | Cameroon                 |    |  | PL | Poland                                       |    |                          |
| CN | China                    | KR | Republic of Korea                        | PT | Portugal                                     |    |                          |
| CU | Cuba                     | KZ | Kazakstan                                | RO | Romania                                      |    |                          |
| CZ | Czech Republic           | LC | Saint Lucia                              | RU | Russian Federation                           |    |                          |
| DE | Germany                  | LI | Liechtenstein                            | SD | Sudan  |    |                          |
| DK | Denmark                  | LK | Sri Lanka                                | SE | Sweden                                       |    |                          |
| EE | Estonia                  | LR | Liberia                                  | SG | Singapore                                    |    |                          |

## 32 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using secreted proteins or the genes that encode them.

### *Summary of the Invention*

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

### *Detailed Description*

#### **Definitions**

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 10801 University Boulevard,  
5 Manassas, Virginia 20110-2209, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained  
10 in SEQ ID NO:X, the complement thereof, or the cDNA within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the  
15 filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages  
20 of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6X SSPE (20X SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1XSSPE, 0.1% SDS. In addition, to achieve even  
25 lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5X SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include  
30 Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such  
35 as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

5       The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxribonucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and  
10       double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability  
15       or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA; thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

      The polypeptide of the present invention can be composed of amino acids joined  
20       to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs,  
25       as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be  
30       branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslation natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a  
35       nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor-formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., Meth Enzymol 182:626-646 (1990); Rattan et al., Ann NY Acad Sci 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## **Polynucleotides and Polypeptides of the Invention**

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 1**

This gene maps to chromosome 3 and therefore polynucleotides of the present invention can be used in linkage analysis as a marker for chromosome 3.

This gene is expressed in several fetal tissues including brain, liver and lung and to a lesser extent in adult tissues, particularly skin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, a variety of cancers, particularly of the brain, liver, and lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells, particularly of the central nervous system, hepatic system, and hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, liver, lung, and skin, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful as a target for a variety of blocking agents, as they are likely to be involved in the promotion of a variety of cancers.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 2

In specific embodiments, the polypeptides of the invention comprise the sequence: MSVPAFIDISEEDQAAELRAYLKSKGAEISEENSEGGLHVDLAQIEAC DVCLKEDDKDVESVMNSVVSLLLILEPDKQEALIESLCEKLVKFREGERPSLRLQLLSNLFHGMDKNTPVRYTVYCSLIKVAASCGAIQYIPTELDQVRKWISDWNLTTEKKHTLLRLLYEALVDCKKSDAASKVMVELLGSYTEDNASQARVDAHRCIVRALKDPNAFLFDHLLTLKPVKFLEGELIHDLLTIFVSAKLASYVKFYQNNKDFIDSL GLLHEQNMAMRLLTFMGMAVENKEISFDTMQQELQIGADDVEAFVIDAVRTKMVYCKIDQTRKVVVSHSTHRTFGKQQWQQLYDTLNAWKQNLNKVKNLSLSLSDT (SEQ ID NO:85), MSVPAFIDISEED (SEQ ID NO:86), QAAELRAYLKSKGAE (SEQ ID NO:87), IEENSEGGLHVDLAQI (SEQ ID NO:88), IEACDVCLKEDDKDVESV (SEQ ID NO:89), VARPSSLFRSAWSCEW (SEQ ID NO:90), LRLQLLSNLFHG (SEQ ID NO:91), KDVESVMNSVVSLLLIL (SEQ ID NO:92), DAASKVMVELLGSYTEDNASQARVDA (SEQ ID NO:93), and/or VEAFFVIDAVR (SEQ ID NO:94). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed in bone and to a lesser extent in brain, lung, T-cells, muscle, skin, testis, spleen and macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, bone cancer, osteoarthritis, and autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

number of disorders of the above tissues or cells; particularly of the immune system and skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, T-cells and other cells and tissue of the immune system, lung, muscle, 5 skin, and testis and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those 10 comprising a sequence shown in SEQ ID NO:49 as residues: Arg-31 to Ser-37, Met-50 to Val-56, Glu-80 to Trp-87, Thr-94 to His-99, Tyr-129 to Ser-135, Tyr-193 to Phe-199, Ser-274 to Gln-285, and/or Ala-293 to Lys-302.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 3

15 The translation product of this gene shares sequence homology with various kinases. The closest homolog is mouse TIF1 which is a mouse nuclear protein. TIF1 enhances RXR and RAR AF-2 in yeast and interacts in a ligand-dependent manner with several nuclear receptors in yeast and mammalian cells, as well as in vitro. Remarkably, these interactions require the amino acids constituting the AF-2 activating domain 20 conserved in all active NRs. Moreover, the oestrogen receptor (ER) AF-2 antagonist hydroxytamoxifen cannot promote ER-TIF1 interaction. We propose that TIF1, which contains several conserved domains found in transcriptional regulatory proteins, is a mediator of ligand-dependent AF-2. Interestingly, the TIF1 N-terminal moiety is fused to B-raf in the mouse oncoprotein T18.

25 This gene is expressed primarily in activated T-cells and to a lesser extent in various other tissues including testes and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are 30 not limited to, autoimmune diseases, AIDS, leukemias, and various other cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be 35 routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, testes and other reproductive tissue, and brain and other

tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:50 as residues: Ala-31 to Glu-36.

The tissue distribution and homology to TIF indicates that polynucleotides and polypeptides corresponding to this gene are useful for modulation of nuclear receptor and ligand interaction in various immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene maps to chromosome 11. Accordingly, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 11. In specific embodiments, the polypeptides of the invention comprise the sequence: MSEIYLRCQDEQQYARWMAGCRLASKGRTMADSSY (SEQ ID NO:95), LVAPRFQRKFKAKQLTPRILEAHQNVAQLSLAEAQLRFIQAWQSL (SEQ ID NO:96), VGDVVKTWRFNSNMRQWNVNWDIR (SEQ ID NO:97), EEIDCTEEEMMVFAALQYHINKLSQS (SEQ ID NO:98), and/or EEIDCTEEEMMVFAALQYHINKLSQS (SEQ ID NO:99). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is expressed primarily in several white blood cell types including monocytes, T-cells, and neutrophils and to a lesser extent in a limited number of other tissues including umbilical vein and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, various diseases of the immune system including AIDS, immunodeficiency diseases, and autoimmune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily

fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:51 as residues: Ser-3 to Pro-9, Leu-17 to Leu-29, Asp-64 to Pro-69, Ile-105 to Gln-110, Thr-183 to Gln-200, Cys-239 to Arg-247, Ser-256 to Met-261, Gln-280 to Ala-296, Arg-310 to Thr-321, Lys-363 to Asp-368, Ser-395 to Trp-400, and/or Thr-443 to Asp-453.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for replacement therapy in a variety of immune system disorders.

#### 10 **FEATURES OF PROTEIN ENCODED BY GENE NO: 5**

This gene is expressed primarily in brain and little or not at all in any other tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, mood disorders, schizophrenia and related diseases, bipolar disorder and unipolar depression. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:52 as residues: Met-1 to Gly-8, Pro-10 to Arg-17, Pro-45 to Ser-55, and/or Gly-63 to Tyr-74.

The tissue distribution of this gene primarily in brain indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder. Also given the brain-specific expression of this gene, the promoter region of this gene contains a brain-specific element that could be used for targeting expression of vector systems to the brain in gene replacement therapy.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 6**

This gene maps to chromosome 1 and therefore, polynucleotides of the invention can be used in linkage analysis as a marker for chromosome 1.

5 This gene is expressed abundantly in rhabdomyosarcoma, is expressed to a high level and in different regions of the brain and pituitary gland and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological disorders and muscular disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., smooth muscle, brain and other tissue of the nervous system, and pituitary, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The abundant expression of this gene in rhabdomyosarcoma indicates a role for the protein product either in the detection and/or treatment of skeletal muscle disorders including muscle degeneration, muscle wasting, and rhabdomyolysis. Furthermore expression in the brain indicates a role for the protein product of this gene in the detection/treatment of neurodegenerative disease states and behavioral disorders such as Alzheimer's Disease, Parkinson's Disease, Huntingtons Disease, schizophrenia, mania, dementia, paranoia, obsessive compulsive disorder and panic disorder.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 7**

30 The translation product of this gene shares sequence homology with the TDAG51 gene which is thought to be important in the mediation of apoptosis and cell death by coupling TCR stimulation to Fas expression. In specific embodiments, the polypeptides of the invention comprise the sequence: KELSFAKAVECVSTGR HIYFTLV(SEQ ID NO:100) and/or GWNAQITLGLVKFKNQQ (SEQ ID NO:101).

35 This gene is expressed in various human tissues including macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., macrophages and other blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:54 as residues: Met-1 to Pro-9, Gln-43 to Glu-49, and/or Phe-95 to Arg-102.

The tissue distribution and homology to TDAG51 gene indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and intervention of immune disorders, such as immunodeficiency, allergy, infection, inflammation, tissue/organ transplantation.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 8**

This gene is expressed in breast tissue, and amniotic cells and to a lesser extent in smooth muscle, T-cells, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fetal distress syndrome and embryonic wasting. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the female reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, amniotic cells, smooth muscle, brain and other tissue of the nervous system, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

5 In specific embodiments, the polypeptides of the invention comprise the sequence: LVLGLSXLNNSYNFSF (SEQ ID NO:102), HVVIGSQAEEGQYSLNF (SEQ ID NO:103), HNCNNSVPGKEHPFDITVM (SEQ ID NO:104), FIKYVLSD KEKKVFGIV (SEQ ID NO:105), IPMQVLANVAYII (SEQ ID NO:106), IPMQVL ANVAYII (SEQ ID NO:107), DGKVAVNLA KLKLFR (SEQ ID NO:108), and/or  
10 IREKNPDGFLSAA (SEQ ID NO:109). Polynucleotides encoding these polypeptides are also encompassed by the invention.

This gene is primarily expressed in the fetal liver, spleen and pituitary gland, and to a lesser extent in multiple tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as  
15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune disorders and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of  
20 the above tissues or cells, particularly of the hepatic, immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., liver, spleen, and pituitary gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,  
25 relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:56 as residues: Ser-62 to Cys-71, Thr-78 to Leu-86, Ser-104 to Lys-109, Ser-130 to Ala-135, and/or Gln-168 to Asp-174.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hepatic disorders, and disorders of the immune and hematopoietic systems, such as hepatic failure, hepatitis, alcoholic liver diseases, portal hypertension, toxic liver injury, liver transplantation, and neoplasm of the liver. The expression in the fetal liver spleen also  
35 indicates its function in hematopoiesis, and therefore the gene may be useful in hematopoietic disorders including anemia, leukemia or cancer

radiotherapy/chemotherapy. The expression in the pituitary gland may indicate its use in endocrine disorders with systemic or specific manifestations.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 10

5 The translation product of this gene shares sequence homology with a chicken DNA binding protein which is thought to be important in transcriptional regulation of gene expression. In specific embodiments, polypeptides of the invention comprise the sequence: MMFGGYETI (SEQ ID NO:110), YRDESSSELSVDSEVEFQLYSQIH (SEQ ID NO:111), YAQDLDDVIREEEHEEKNSGNSSESSSSKPNQKKLIVLSDSEVI  
 10 QLSDGSEVITLSDSDSIYRCKGKNVRVQAQENAHGLSSSLQSNELVDKKCKSDI  
 EKPKEERSGVIREVMIEVSSSEEEESTISEGDNVESW (SEQ ID NO:112), MLLG  
 CEVDDKDDDDILLNLVGCENSVTEGEDGINWSIS (SEQ ID NO:113), DKDIEAQI  
 ANNRTPGRWT (SEQ ID NO:114), QRYYSANKNIICRNCDKRGHLSKNCPLP  
 RKV (SEQ ID NO:115), and/or RRCFLCSRGRHLLYSCPAPLCEYCPVPKMLDHS  
 15 CLFRHSWDKQCDRCHMLGHYTDACTEIWRQYHLTTKPGPPKKPKTPSRPSAL  
 AYC YHCAQKGHYGHECPEREVYDPSVSPFICYXDKYEIQEREKRLKQKIKV  
 XKKNGVIEPEPSKLPYIKAANENPHHDIRKGRASWKS NRWPQ (SEQ ID  
 NO:116). Polynucleotides encoding these polypeptides are also encompassed by the invention.

20 This gene is expressed in tonsils and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the immune, hematopoietic, and lymphatic systems.

25 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, hematopoietic, and lymph systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., tonsils, and  
 30 bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to DNA binding protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of disorders in the immune, hematopoietic, and lymph systems.

## 5    **FEATURES OF PROTEIN ENCODED BY GENE NO: 11**

This gene is expressed in dendritic and T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cells types (e.g., dendritic cells, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the treatment and diagnosis of immune system disorders, particularly those involving dendritic or T-cells such as inflammation.

## 25    **FEATURES OF PROTEIN ENCODED BY GENE NO: 12**

This gene is expressed in activated neutrophils, endothelial cells, T cells and to a lesser extent in brain and liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, AIDS, immune disorders and susceptibility to infectious disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neutrophils and other

blood cells, endothelial cells, T-cells and other cells and tissue of the immune system, brain and other tissue of the nervous system, and liver, and cancerous and wounded tissues)-or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:59 as residues: Glu-41 to Val-46.

This gene product is useful for the diagnosis and/or treatment of a variety of disorders, including hematopoietic disorders, neurological disorders, liver disease, and disorders involving angiogenesis.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene is expressed in keratinocytes and to a lesser extent in endothelial cells and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, impaired wound healing. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cells types (e.g., keratinocytes and other cells of the skin, endothelial cells, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:60 as residues: Pro-35 to Trp-42, Ala-53 to Asp-62, and/or Arg-103 to Pro-113.

The tissue distribution indicates that the protein products of this gene are useful for the treatment of wound healing deficiency and skin disorders.

### FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene is expressed in kidney and to a lesser extent in embryonic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions which include, but are not limited to, renal failure. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., kidney, embryonic and other rapidly developing (e.g., dividing) tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 15**

This gene is expressed primarily in brain and to a lesser extent in liver. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, depression, manic depression and other mental diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., brain and other tissue of the nervous system, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the treatment of central nervous system disorders such as depression and other mental illnesses.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 16**

This gene is expressed in fetal brain and to a lesser extent in placenta, endothelial cells, fetal lung, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, restinosis, birth defects and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cardiovascular system, and developmental process, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, lung, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:63 as residues: Gln-36 to Lys-42, and/or Glu-89 to Arg-104.

The tissue distribution indicates that the protein products of this gene are useful for the development of agonists and/or antagonists for treatment of nervous system disorders and fetal development.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 17**

This gene is expressed in hemangiopericytoma and to a lesser extent in fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, hemangiopericytomas and other cancers, as well as developmental disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., vascular tissue, pericytic tissue, and developing tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include

those comprising a sequence shown in SEQ ID NO:64 as residues: Glu-43 to Pro-51, Gly-71 to Arg-82, Pro-96 to Arg-103, and/or Thr-130 to Gly-140.

The polynucleotides and polypeptides related to this gene are believed to be useful for the treatment and diagnosis of tumors, particularly hemangiopericytomas, and for the treatment of developmental disorders.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 18**

This gene is expressed in fetal liver and to a lesser extent in brain and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, fetal disorders, fetal development, and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, brain and other tissue of the nervous system, and T-cells and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the identification of agonists and /or antagonists for treatment of mental illnesses such as schizophrenia and depression. The gene product may also be useful for monitoring fetal development during pregnancy.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 19**

This gene is expressed in T cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, central nervous diseases and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of

disorders of the above tissues or cells, particularly of the central nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:66 as residues:  
Lys-69 to Leu-74, Ser-92 to Phe-97, Asp-109 to Leu-117, Leu-142 to Ser-159, Thr-166 to Glu-183, Ala-191 to Glu-205, and/or Pro-213 to Glu-220.

The tissue distribution indicates that the protein products of this gene are useful for the development of drugs for treatment of disorders affecting the central nervous system and immune system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 20

The translation product of this gene shares sequence homology with a *C. elegans* ORF that seems to be a transmembrane protein. (See GenBank Accession No. 790406.) This contig has two probable frameshifts between the +2 and +3 frames based on homology with the *C. elegans* gene. This frameshift can easily be resolved by sequencing the deposited clone. Moreover, this gene maps to chromosome 8, and therefore can be used as a marker in linkage analysis for chromosome 8.

This gene is expressed ubiquitously, including T cells and amygdala.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, amygdala, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The ubiquitous tissue distribution and homology to a *C. elegans* transmembrane-like protein indicates that the protein product of this gene plays a role important in both vertebrates and invertebrates and is useful for diagnosis or treatment of disorders related to this gene.

5

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 21**

This gene is expressed primarily in embryonic and testes and to a lesser extent in ovary, hepatoma, kidney, endothelial, and smooth muscle cells.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, metabolic disorder, abnormal embryonic development and tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell  
15 type(s). For a number of disorders of the above tissues or cells, particularly of the embryonic or vascular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, kidney, endothelial cells, and smooth muscle cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or  
20 spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to NADH dehydrogenase indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis  
25 and/or treating metabolic disorders, particularly involving embryonic and vascular tissues.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 22**

The translation product of this gene shares sequence homology with alpha 1C  
30 adrenergic receptor which is thought to be important in neuronal signal transmission.

This gene is expressed primarily in breast lymphnode and to a lesser extent in uterine cancer and testis tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
35 biological sample and for diagnosis of diseases and conditions which include, but are not limited to, neurological disorders. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neurologic, breast lymphonode, uterine cancer, and testis, expression of this gene at significantly higher or lower levels may be routinely  
5 detected in certain tissues (e.g., breast tissue, lymphoid tissue, uterine tissue, and testis and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an  
10 individual not having the disorder.

The tissue distribution and homology to alpha 1C adrenergic receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful for transmitting signals to neurons.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 23**

The translation product of this gene shares sequence homology with G-protein-coupled receptor which is thought to be important in mediating a wide variety of physiological function and belongs to a gene superfamily with members ranging from chemokine receptor to bradykinin receptor. This gene has also recently been cloned by another  
20 group, calling the gene platelet activating receptor homolog. (See GenBank Accession No. 2580588.) Preferred polypeptide fragments comprise the amino acid sequence: LSIIFLAFVSDRCLQL (SEQ ID NO:117) and GSCFATWAFIQKNTNHRCSVSIY LINLLTADFLTLALPVKIVVDLGVAPWKLKIFHCQVTACLIYIN (SEQ ID NO:118). Also preferred are polynucleotide fragments encoding these polypeptide  
25 fragments.

This gene is expressed primarily in immune cells, particularly lymphocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are  
30 not limited to, disorders of lymphocytes and other immune cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected  
35 in certain tissues and cell types (e.g., lymphocytes and other cells and tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum,

plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID  
5 NO:70 as residues: Asp-59 to Asn-65, Lys-72 to Trp-79, Tyr-110 to Val-121, and/or Ala-204 to Asn-215.

The tissue distribution and homology to G-protein coupled receptor indicates that polynucleotides and polypeptides corresponding to this gene are useful as  
10 chemokine receptor on lymphocytes that regulate immune response.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 24**

The translation product of this gene shares sequence homology with protein disulfide isomerase which is thought to be important in protein folding and protein-protein interaction. This gene also shares homology to genes having thioredoxin  
15 domains. (See Accession No. 1943817.) This gene also maps to chromosome 9, and therefore may be useful in linkage analysis as a marker for chromosome 9.

This gene is expressed primarily in tumor tissues and to a lesser extent in a wide variety of normal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as  
20 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders due to inappropriate protein folding and protein-protein interaction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s)  
25 or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumorigenic process, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the  
30 standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:71 as residues: Glu-78 to Asn-83, Asp-91 to Gln-100, Glu-122 to Ser-128, Arg-137 to Pro-143, Asp-157 to Asn-162, Glu-168 to Asn-174, Ser-199 to Gly-206, Pro-213 to Ala-218, Glu-251 to Thr-257, Ser-35  
35 353 to His-361, Gly-363 to Ala-375, Pro-382 to Phe-387, and/or Arg-401 to Leu-406.

The tissue distribution and homology to protein disulfide isomerase indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating protein folding and protein-protein interaction in tumor tissues.

#### 5    **FEATURES OF PROTEIN ENCODED BY GENE NO: 25**

This gene is expressed primarily in leukocytes involved in immune defense, including T cells, macrophages, neutrophils and to a lesser extent in synovium, adrenal gland tumor, adipose, and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as  
10 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, defects or disorders in leukocytes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of  
15 the above tissues or cells, particularly of the immune and defense systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., leukocytes and other cells and tissues of the immune system, synovium, adrenal gland, adipose and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or  
20 another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating leukocyte function and may be used  
25 for diagnosis and treatment of disorders in immune and defense systems.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 26**

This gene is expressed in a variety of tissues and cell types, including colon cancer, breast cancer, neutrophils, T cells, spinal cord, fibroblasts, and vascular  
30 endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer, disorder and abnormalities in leukocytes and other tissues.  
35 Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly those cells involved in tumorigenesis and immune defense systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, breast tissue, neutrophils, T-cells and other blood cells, spinal cord and other tissue of the nervous system, endothelial cells, vascular tissue, and fibroblasts, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer or immune system disorders.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 27**

The translation product of this gene shares sequence homology with a mouse pancreatic polypeptide. (See GenBankAccession No. 200464.) Thus, it is likely that this gene has activity similar to the mouse pancreatic polypeptide. Preferred polypeptide fragments comprise the amino acids sequence: APLETMQNKPRAPQKRALPFPEL ELRDYASVLTRYSLGLRNKEPSLGHRWGTQKLGRSPC (SEQ ID NO:119). Also preferred are polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in neutrophils and to a lesser extent in induced endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, disorders in neutrophils or leukocyte adhesion. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., neutrophils and other blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulation of neutrophils or leukocyte adhesion to endothelial cells. It may be used to diagnose or treat disorders associated with neutrophils and vascular endothelial cells.

5

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 28**

This gene is expressed primarily in prostate BPH.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, benign hypertrophy of the prostate. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the male urogenital system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15  
20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of benign hypertrophy of the prostate or prostate cancer.

#### **25 FEATURES OF PROTEIN ENCODED BY GENE NO: 29**

The translation product of this gene shares sequence homology with C16C10.7, a *C. elegans* gene similar to zinc finger protein, a protein involved in DNA binding. Thus, this protein is expected to share certain biological activities with C16C10.7 including DNA binding activities.

30

This gene is expressed primarily in activated T-cells and to a lesser extent in fetal brain, TNF-induced amniotic cells and epididymus.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immune or neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

35

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and central nervous systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, brain and other tissue of the nervous system, amniotic cells, and epididymus and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein products of this gene are useful for the diagnosis and treatment of immune and/or neurodegenerative disorders and promotion of survival and differentiation of neurons.

#### 15 **FEATURES OF PROTEIN ENCODED BY GENE NO: 30**

This gene is expressed primarily in T-cells and to a lesser extent in bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, immunological disorders including autoimmune disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., T-cells and other cells and tissue of the immune system, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. It is believed that this gene maps to chromosome 4: Transcript map: WI-11395, Chr.4, D4S395-D4S414; Whitehead map: WI-11395, Chr.4, 498.0 cR; dbSTS entries: G21269.

The tissue distribution indicates that the protein products of this gene are useful for diagnosis and treatment of immunologically mediated disorders as they are thought

to play a role in the proliferation, survival, differentiation, and/or activation of a variety of hematopoietic cells, including early progenitors or hematopoietic stem cells.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 31**

5        This gene is expressed primarily in human skin.

         Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, wound healing and skin cancers. Similarly, polypeptides and antibodies  
10       directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the integumentary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., skin and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma,  
15       urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

         The tissue distribution indicates that the protein products of this gene are useful  
20       for diagnosis and treatment of skin cancers and wound healing.

#### **FEATURES OF PROTEIN ENCODED BY GENE NO: 32**

         The translation product of this gene shares sequence homology with human Tear Prealbumin (GenBank accession no. gil307518) and rat Oderant-binding protein  
25       (GenBank accession no. gil207551), both of which are thought to be important in molecule binding and transport.

         This gene is expressed primarily in endometrial tumor.

         Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a  
30       biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers of the endometrium, skin and haemopoietic system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the haemopoietic  
35       system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues (e.g., cells and tissue of the immune system, and

endometrium and other tissue of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to the molecule binding and transport gene family indicates that the protein products of this gene are useful for the diagnosis and treatment of cancers of the endometrium and haemopoietic system as well as for the treatment of autoimmune disorders such as inflammation.

10

| Gene No. | cDNA Clone ID | ATCC Deposit Nr and Date | Vector      | NT SEQ ID NO: X | Total NT Seq. | 5' NT of Clone Seq. | 3' NT of Clone Seq. | 5' NT of Start Codon | 5' NT of First AA of Signal Pep | AA SEQ ID NO: Y | First AA of Sig Pep | Last AA of Sig Pep | First AA of Secreted Portion | Last AA of ORF  |
|----------|---------------|--------------------------|-------------|-----------------|---------------|---------------------|---------------------|----------------------|---------------------------------|-----------------|---------------------|--------------------|------------------------------|-----------------|
| 1        | HSVVBZ80      | 209075<br>05/22/97       | Uni-ZAP XR  | 11              | 1169          | 64                  | 1060                | 162                  | 162                             | 48              | 1                   | 38                 | 39                           | 145             |
| 2        | HTAAU21       | 209075<br>05/22/97       | Uni-ZAP XR  | 12              | 1310          | 1                   | 1310                | 283                  | 283                             | 49              | 1                   | 18                 | 19                           | 311             |
| 3        | HTLEK16       | 209075<br>05/22/97       | Uni-ZAP XR  | 13              | 1139          | 19                  | 1111                |                      | 251                             | 50              | 1                   | 21                 | 22                           | 46              |
| 4        | HUSIR91       | 209075<br>05/22/97       | pSport1     | 14              | 2271          | 743                 | 2271                | 59                   | 59                              | 51              | 1                   | 23                 | 24                           | 466             |
| 4        | HUSIR91       | 209075<br>05/22/97       | pSport1     | 43              | 2581          | 1035                | 2164                | 1148                 | 1148                            | 80              | 1                   | 27                 | 28                           | 207             |
| 5        | HADMC21       | 209075<br>05/22/97       | pBluescript | 15              | 626           | 60                  | 479                 | 91                   | 91                              | 52              | 1                   | 51                 | 52                           | 82              |
| 6        | HAGFM45       | 209075<br>05/22/97       | Uni-ZAP XR  | 16              | 2118          | 1170                | 2058                | 1248                 | 1248                            | 53              | 1                   | 16                 | 17                           | 62              |
| 7        | HAIBE65       | 209075<br>05/22/97       | Uni-ZAP XR  | 17              | 1076          | 396                 | 993                 | 528                  | 528                             | 54              | 1                   | 31                 | 32                           | 123             |
| 8        | HAQBH57       | 209075<br>05/22/97       | Uni-ZAP XR  | 18              | 1379          | 420                 | 1306                | 618                  | 618                             | 55              | 1                   | 25                 | 26                           | 179             |
| 9        | HATCX80       | 209075<br>05/22/97       | Uni-ZAP XR  | 19              | 1337          | 47                  | 1337                | 199                  | 199                             | 56              | 1                   | 18                 | 19                           | 286             |
| 10       | HCFLQ84       | 209075<br>05/22/97       | pSport1     | 20              | 1390          | 237                 | 1390                | 410                  | 410                             | 57              | 1                   | 20                 | 21                           | 33 <sup>2</sup> |
| 11       | HCFLS78       | 209075<br>05/22/97       | pSport1     | 21              | 1431          | 178                 | 981                 | 420                  | 420                             | 58              | 1                   | 21                 | 22                           | 23              |

| Gene No. | cDNA Clone ID | ATCC Deposit Nr and Date | Vector        | NT SEQ ID NO: X | Total NT Seq. | 5' NT of Clone Seq. | 3' NT of Clone Seq. | 5' NT of Start Codon | 5' NT of First AA of Signal Pep | AA SEQ ID NO: Y | First AA of Sig Pep | Last AA of Sig Pep | First AA of Secreted Portion | Last AA of ORF |
|----------|---------------|--------------------------|---------------|-----------------|---------------|---------------------|---------------------|----------------------|---------------------------------|-----------------|---------------------|--------------------|------------------------------|----------------|
| 12       | HTADI12       | 209075<br>05/22/97       | Uni-ZAP XR    | 22              | 2539          | 69                  | 2539                | 104                  | 104                             | 59              | 1                   | 27                 | 28                           | 46             |
| 13       | HEMCM42       | 209075<br>05/22/97       | Uni-ZAP XR    | 23              | 1041          | 48                  | 1007                | 58                   | 58                              | 60              | 1                   | 29                 | 30                           | 113            |
| 14       | HEONP72       | 209075<br>05/22/97       | pSport1       | 24              | 1962          | 1                   | 1947                | 181                  | 181                             | 61              | 1                   | 19                 | 20                           | 31             |
| 15       | HFCDW34       | 209075<br>05/22/97       | Uni-ZAP XR    | 25              | 1228          | 321                 | 1228                | 525                  | 525                             | 62              | 1                   | 24                 | 25                           | 80             |
| 16       | HTTEU91       | 209075<br>05/22/97       | Uni-ZAP XR    | 26              | 1340          | 325                 | 1340                | 15                   | 15                              | 63              | 1                   | 18                 | 19                           | 103            |
| 17       | HHGBF89       | 209075<br>05/22/97       | Lambda ZAP II | 27              | 806           | 31                  | 806                 | 77                   | 77                              | 64              | 1                   | 19                 | 20                           | 145            |
| 17       | HHGBF89       | 209075<br>05/22/97       | Lambda ZAP II | 45              | 796           | 31                  | 796                 | 77                   | 77                              | 82              | 1                   | 25                 | 26                           | 145            |
| 18       | HKIYQ65       | 209075<br>05/22/97       | pBluescript   | 28              | 696           | 1                   | 684                 | 98                   | 98                              | 65              | 1                   | 17                 | 18                           | 30             |
| 19       | HKMLN27       | 209075<br>05/22/97       | pBluescript   | 29              | 1007          | 71                  | 963                 | 129                  | 129                             | 66              | 1                   | 23                 | 24                           | 259            |
| 20       | HKIAC30       | 209022<br>05/08/97       | Uni-ZAP XR    | 30              | 2017          | 126                 | 2007                | 161                  | 161                             | 67              | 1                   |                    |                              | 22             |
| 21       | HKIXB95       | 209022<br>05/08/97       | pBluescript   | 31              | 699           | 196                 | 699                 | 230                  | 230                             | 68              | 1                   | 22                 | 23                           | 26             |
| 22       | HLMY86        | 209022<br>05/08/97       | Lambda ZAP II | 32              | 1264          | 1                   | 1264                |                      | 342                             | 69              | 1                   | 16                 | 17                           | 28             |
| 23       | HLYAZ61       | 209022<br>05/08/97       | pSport1       | 33              | 997           | 74                  | 997                 | 205                  | 205                             | 70              | 1                   | 20                 | 21                           | 215            |

| Gene No. | cDNA Clone ID | ATCC Deposit Nr and Date | Vector     | NT SEQ ID NO: X | Total NT Seq. | 5' NT of Clone Seq. | 3' NT of Clone Seq. | 5' NT of Start Codon | 5' NT of First AA of Signal Pep | AA SEQ ID NO: Y | First AA of Sig Pep | Last AA of Sig Pep | First AA of Secreted Portion | Last AA of ORF |
|----------|---------------|--------------------------|------------|-----------------|---------------|---------------------|---------------------|----------------------|---------------------------------|-----------------|---------------------|--------------------|------------------------------|----------------|
| 24       | HMQDT36       | 209022<br>05/08/97       | Uni-ZAP XR | 34              | 1914          | 37                  | 1897                | 192                  | 192                             | 71              | 1                   | 32                 | 33                           | 406            |
| 25       | HNEDF25       | 209022<br>05/08/97       | Uni-ZAP XR | 35              | 1020          | 11                  | 1020                |                      | 211                             | 72              | 1                   |                    |                              | 8              |
| 26       | HNFET17       | 209022<br>05/08/97       | Uni-ZAP XR | 36              | 781           | 31                  | 781                 | 100                  | 100                             | 73              | 1                   |                    |                              | 33             |
| 27       | HNHCR46       | 209022<br>05/08/97       | Uni-ZAP XR | 37              | 966           | 507                 | 948                 |                      | 576                             | 74              | 1                   | 29                 | 30                           | 56             |
| 28       | HPWAS91       | 209022<br>05/08/97       | Uni-ZAP XR | 38              | 416           | 1                   | 416                 | 95                   | 95                              | 75              | 1                   | 24                 | 25                           | 25             |
| 29       | HWTAW41       | 209022<br>05/08/97       | Uni-ZAP XR | 39              | 1114          | 804                 | 1114                | 843                  | 843                             | 76              | 1                   |                    |                              | 14             |
| 30       | HBMUT52       | 209022<br>05/08/97       | Uni-ZAP XR | 40              | 602           | 142                 | 602                 | 204                  | 204                             | 77              | 1                   | 26                 | 27                           | 32             |
| 31       | HERAG83       | 209022<br>05/08/97       | Uni-ZAP XR | 41              | 970           | 1                   | 970                 | 110                  | 110                             | 78              | 1                   | 22                 | 23                           | 22             |
| 32       | HETFI51       | 209022<br>05/08/97       | Uni-ZAP XR | 42              | 1002          | 1                   | 1002                | 43                   | 43                              | 79              | 1                   | 21                 | 22                           | 172            |
| 32       | HETFI51       | 209022<br>05/08/97       | Uni-ZAP XR | 47              | 981           | 1                   | 981                 | 23                   | 23                              | 84              | 1                   | 17                 | 18                           | 30             |

Table 1 summarizes the information corresponding to each "Gene No." described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may

be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below). It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

#### 15 Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinje, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinje, *supra*.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinje are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in

some cases, cleavage of the signal sequence from a secreted protein is not entirely uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

### **Polynucleotide and Polypeptide Variants**

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

By a polynucleotide having a nucleotide sequence at least, for example, 95% "identical" to a reference nucleotide sequence of the present invention, it is intended that the nucleotide sequence of the polynucleotide is identical to the reference sequence except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the reference nucleotide sequence encoding the polypeptide. In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to a reference nucleotide sequence, up to 5% of the nucleotides in the reference sequence may be deleted or substituted with another nucleotide, or a number of nucleotides up to 5% of the total nucleotides in the reference sequence may be inserted into the reference sequence. The query sequence may be an entire sequence shown in Table 1, the ORF (open reading frame), or any fragment specified as described herein.

As a practical matter, whether any particular nucleic acid molecule or polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to a nucleotide sequence of the present invention can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are both DNA sequences. An RNA sequence can be compared by converting U's to T's. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB alignment of DNA sequences to calculate percent identity are:

Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, Window Size=500 or the length of the subject nucleotide sequence, whichever is shorter.

5 If the subject sequence is shorter than the query sequence because of 5' or 3' deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for 5' and 3' truncations of the subject sequence when calculating percent identity. For subject sequences truncated at the 5' or 3' ends, relative to the query sequence, the percent identity is corrected by calculating the number of bases of the query sequence that are 5' and 3' of the subject sequence, which are not matched/aligned, as a percent of the total bases of the query sequence. Whether a nucleotide is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This corrected score is what is used for the purposes of the present invention. Only bases outside the 5' and 3' bases of the subject sequence, as displayed by the FASTDB alignment, which are not matched/aligned with the query sequence, are calculated for the purposes of manually adjusting the percent identity score.

For example, a 90 base subject sequence is aligned to a 100 base query sequence to determine percent identity. The deletions occur at the 5' end of the subject sequence and therefore, the FASTDB alignment does not show a matched/alignment of the first 10 bases at 5' end. The 10 unpaired bases represent 10% of the sequence (number of bases at the 5' and 3' ends not matched/total number of bases in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 bases were perfectly matched the final percent identity would be 90%. In another example, a 90 base subject sequence is compared with a 100 base query sequence. This time the deletions are internal deletions so that there are no bases on the 5' or 3' of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only bases 5' and 3' of the subject sequence which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

By a polypeptide having an amino acid sequence at least, for example, 95% "identical" to a query amino acid sequence of the present invention, it is intended that the amino acid sequence of the subject polypeptide is identical to the query sequence except that the subject polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the query amino acid sequence. In other words,

to obtain a polypeptide having an amino acid sequence at least 95% identical to a query amino acid sequence, up to 5% of the amino acid residues in the subject sequence may be inserted, deleted, (indels) or substituted with another amino acid. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

As a practical matter, whether any particular polypeptide is at least 90%, 95%, 96%, 97%, 98% or 99% identical to, for instance, the amino acid sequences shown in Table 1 or to the amino acid sequence encoded by deposited DNA clone can be determined conventionally using known computer programs. A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag et al. (Comp. App. Biosci. (1990) 6:237-245). In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB amino acid alignment are: Matrix=PAM 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject amino acid sequence, whichever is shorter.

If the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction must be made to the results. This is because the FASTDB program does not account for N- and C-terminal truncations of the subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity is corrected by calculating the number of residues of the query sequence that are N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. Whether a residue is matched/aligned is determined by results of the FASTDB sequence alignment. This percentage is then subtracted from the percent identity, calculated by the above FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score is what is used for the purposes of the present invention. Only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the

subject sequence.

For example, a 90 amino acid residue subject sequence is aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C- termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity would be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for. No other manual corrections are to be made for the purposes of the present invention.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

Using known methods of protein engineering and recombinant DNA technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988

(1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., J. Biotechnology 7:199-216 (1988).)

5        Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (J. Biol. Chem 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over 10 the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild- 15 type.

      Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form 20 will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

25        Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., 30 Science 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

      The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino 35 acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these

positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, Science 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln, replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

### **Polynucleotide and Polypeptide Fragments**

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, 701-750, 751-800, 800-850, 851-900, 901-950, 951-1000, 1001-1050, 1051-1100, 1101-1150, 1151-1200, 1201-1250, 1251-1300, 1301-1350, 1351-1400, 1401-1450, 1451-1500, 1501-1550, 1551-1600, 1601-1650, 1651-1700, 1701-1750, 1751-1800, 1801-1850, 1851-1900, 1901-1950, 1951-2000, or 2001 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity. More preferably, these polynucleotides can be used as probes or primers as discussed herein.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, or 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or

the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions.

Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### **Epitopes & Antibodies**

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998- 4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including

monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., Cell 37:767-778 (1984); Sutcliffe, J. G. et al., Science 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., supra; Wilson et al., supra; Chow, M. et al., Proc. Natl. Acad. Sci. USA 82:910-914; and Bittle, F. J. et al., J. Gen. Virol. 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., J. Nucl. Med. 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library. Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### **Fusion Proteins**

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the

polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Traunecker et al., *Nature* 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., *J. Biochem.* 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., *J. Molecular Recognition* 8:52-58 (1995); K. Johanson et al., *J. Biol. Chem.* 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., *Proc. Natl. Acad. Sci. USA* 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein. Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., *Cell* 37:767 (1984).)

Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the present invention.

### **Vectors, Host Cells, and Protein Production**

5       The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

10       The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged in vitro using an appropriate packaging cell line and then transduced into host cells.

15       The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination,  
20       and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

      As indicated, the expression vectors will preferably include at least one  
25       selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect  
30       cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS, 293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

      Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A,  
35       pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1

and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host-cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

#### **Uses of the Polynucleotides**

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers,

since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

5 Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

10 Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include in situ hybridization, prescreening with labeled flow-sorted chromosomes, and preselection by hybridization to construct chromosome  
15 specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence in situ hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides  
20 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for  
25 marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the  
30 physical position of the polynucleotide can be used in linkage analysis. Linkage analysis establishes coinheritance between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library) .) Assuming 1 megabase mapping resolution and one gene per  
35 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinheritance is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined. First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural  
5 alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic  
10 polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic  
15 marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the  
20 region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991) ) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off  
25 of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One  
30 goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

35 The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In

this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

5 A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell . Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene  
10 expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (125I, 121I), carbon (14C), sulfur (35S), tritium (3H), indium (112In), and technetium (99mTc), and fluorescent labels, such as fluorescein and rhodamine, and  
15 biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit  
20 detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hybridoma.

A protein-specific antibody or antibody fragment which has been labeled with  
25 an appropriate detectable imaging moiety, such as a radioisotope (for example, 131I, 112In, 99mTc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety  
30 needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20 millicuries of 99mTc. The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of  
35 Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention can be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

#### **Biological Activities**

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

#### **Immune Activity**

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the

proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, Digeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, antiphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1.)

#### **Hyperproliferative Disorders**

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect

interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenström's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### **Infectious Disease**

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes

Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., Corynebacterium, Mycobacterium, Norcardia), Aspergillosis, Bacillaceae (e.g., Anthrax, Clostridium), Bacteroidaceae, Blastomycosis, Bordetella, Borrelia, Brucellosis, Candidiasis, Campylobacter, Coccidioidomycosis, Cryptococcosis, Dermatocycoses, Enterobacteriaceae (Klebsiella, Salmonella, Serratia, Yersinia), Erysipelothrix, Helicobacter, Legionellosis, Leptospirosis, Listeria, Mycoplasmatales, Neisseriaceae (e.g., Acinetobacter, Gonorrhea, Meningococcal), Pasteurellaceae Infections (e.g., Actinobacillus, Hemophilus, Pasteurella), Pseudomonas, Rickettsiaceae, Chlamydiaceae, Syphilis, and Staphylococcal. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria, Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonias. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### **Regeneration**

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue

regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate  
5 nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral  
10 neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

#### 15 **Chemotaxis**

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of  
20 hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system  
25 disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present  
30 invention may inhibit chemotactic activity. These molecules could also be used to treat disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

#### **Binding Activity**

35 A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit

(antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2):Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

#### **Other Activities**

10 A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

15 A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

20 A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, cardiac rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

25 A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

#### **Other Preferred Embodiments**

30 Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

35 Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of

positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is a nucleic acid molecule wherein said sequence of contiguous  
5 nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Similarly preferred is a nucleic acid molecule wherein said sequence of  
10 contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide  
15 sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Further preferred is an isolated nucleic acid molecule comprising a nucleotide  
20 sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

A further preferred embodiment is a nucleic acid molecule comprising a  
nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ  
ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the  
First Amino Acid of the Signal Peptide and ending with the nucleotide at about the  
25 position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in  
Table 1.

A further preferred embodiment is an isolated nucleic acid molecule comprising  
a nucleotide sequence which is at least 95% identical to the complete nucleotide  
sequence of SEQ ID NO:X.

Also preferred is an isolated nucleic acid molecule which hybridizes under  
30 stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which  
35 comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1,  
which DNA molecule is contained in the material deposited with the American Type

Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide  
5 comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

10 Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

15 Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid  
20 sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1; and an amino acid sequence of a secreted portion of a protein encoded by a human  
25 cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an  
30 individual a pharmaceutical composition comprising an amount of an isolated polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of  
35 illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

5 Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For  
10 example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

|    | <u>Vector Used to Construct Library</u> | <u>Corresponding Deposited Plasmid</u> |
|----|---|--|
|    | Lambda Zap                              | pBluescript (pBS)                      |
|    | Uni-Zap XR                              | pBluescript (pBS)                      |
| 15 | Zap Express                             | pBK                                    |
|    | lafmid BA                               | plafmid BA                             |
|    | pSport1                                 | pSport1                                |
|    | pCMVSPORT 2.0                           | pCMVSPORT 2.0                          |
|    | pCMVSPORT 3.0                           | pCMVSPORT 3.0                          |
| 20 | pCR <sup>®</sup> 2.1                    | pCR <sup>®</sup> 2.1                   |

Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128, 256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 25: 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

35 Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain

DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., *Focus* 15:59 (1993).) Vector lafmid BA (Bento Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into *E. coli* strain XL-1 Blue. Vector pCR<sup>®</sup>2.1, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into *E. coli* strain DH10B, available from Life Technologies. (See, for instance, Clark, J. M., *Nuc. Acids Res.* 16:9677-9686 (1988) and Mead, D. et al., *Bio/Technology* 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported. The oligonucleotide is labeled, for instance, with <sup>32</sup>P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25  $\mu$ l of reaction mixture with 0.5  $\mu$ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM  $MgCl_2$ , 0.01% (w/v) gelatin, 20  $\mu$ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., *Nucleic Acids Res.* 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is

used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

5

**Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic P1 library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

10

**Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with  $P^{32}$  using the rediprime™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

15

20

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

25

**Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on

30

35

either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

5 **Example 5: Bacterial Expression of a Polypeptide**

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as  
10 BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site  
15 (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses  
20 the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml).  
25 The O/N culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.<sup>600</sup>) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalacto pyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

30 Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from  
35 QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high

affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

5 Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

10 The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4° C or frozen at -80° C.

15 In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number 209645, deposited on February 25, 1998.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a T5 phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (*lacIq*). The origin of replication (*oriC*) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

25 DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

30 The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

**Example 6: Purification of a Polypeptide from an Inclusion Body**

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

5        Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50  
10        mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

      The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by  
15        centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

      The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C  
20        overnight to allow further GuHCl extraction.

      Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing  
25        for 12 hours prior to further purification steps.

      To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive  
30        Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5  $\mu$ g of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

#### **Example 7: Cloning and Expression of a Polypeptide in a Baculovirus**

##### **Expression System**

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak *Drosophila* promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gal" (Life

Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.)

5 After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200  $\mu$ l of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested  
10 and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is  
15 removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5  $\mu$ Ci of  $^{35}$ S-methionine and 5  $\mu$ Ci  $^{35}$ S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE  
20 followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

#### **Example 8: Expression of a Polypeptide in Mammalian Cells**

25 The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by  
30 donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from

Retroviruses, e.g., RSV, HTLV, HIV and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden),  
5 pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QC1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

10 Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

The transfected gene can also be amplified to express large amounts of the  
15 encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is  
20 the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992). Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the  
25 production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the  
30 CMV-enhancer (Boshart et al., Cell 41:521-530 (1985).) Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

35 Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five  $\mu$ g of the expression plasmid pC6 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

### **Example 9: Protein Fusions**

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Traunecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the halflife time in vivo. Nuclear localization signals fused to the

polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

#### Human IgG Fc region:

```
GGGATCCGGAGCCCAAATCTTCTGACAAACTCACACATGCCCACCGTGCC
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCCTCTTCCCCC AAAACC
CAAGGACACCCTCATGATCTCCCGGACTCCTGAGGTCACATGCGTGGTGGT
GGACGTAAGCCACGAAGACCCTGAGGTCAAGTTCAACTGGTACGTGGACG
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGCGGGAGGAGCAGTACAAC
AGCACGTACCGTGTGGTCAGCGTCCTACCGTCCTGCACCAGGACTGGCTG
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAAGCCCTCCCAACCCCC
ATCGAGAAAACCATCTCCAAAGCCAAAGGGCAGCCCCGAGAACCACAGGT
GTACACCCTGCCCCCATCCCGGGATGAGCTGACCAAGAACCAGGTCAGCCT
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGGA
GAGCAATGGGCAGCCGGAGAACAACTACAAGACCACGCCTCCCGTGCTGG
ACTCCGACGGCTCCTTCTTCCTCTACAGCAAGCTCACCGTGGACAAGAGCA
```

GGTGGCAGCAGGGGAACGTCTTCTCATGCTCCGTGATGCATGAGGCTCTGC  
ACAACCACTACACGCAGAAGAGCCTCTCCCTGTCTCCGGGTAAATGAGTGC  
GACGGCCGCGACTCTAGAGGAT (SEQ ID NO:1)

5 **Example 10: Production of an Antibody from a Polypeptide**

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with

this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide.

5 Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab')<sub>2</sub> and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such

10 fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab')<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use

15 "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496;

20 Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

#### **Example 11: Production Of Secreted Protein For High-Throughput**

#### **Screening Assays**

25 The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution

30 (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhittaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The

35 PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at  $2 \times 10^5$  cells/well in .5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhittaker))/10% heat inactivated FBS(14-503F Biowhittaker)/1x Penstrep(17-602E Biowhittaker). Let the cells grow overnight.

5 The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31985070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a  
10 multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

15 Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel,  
20 adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (116.6 mg/L of  $\text{CaCl}_2$  (anhyd); 0.00130 mg/L  $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$ ; 0.050 mg/L of  $\text{Fe}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$ ; 0.417 mg/L of  $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ ; 311.80  
25 mg/L of KCl; 28.64 mg/L of  $\text{MgCl}_2$ ; 48.84 mg/L of  $\text{MgSO}_4$ ; 6995.50 mg/L of NaCl; 2400.0 mg/L of  $\text{NaHCO}_3$ ; 62.50 mg/L of  $\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$ ; 71.02 mg/L of  $\text{Na}_2\text{HPO}_4$ ; 4320 mg/L of  $\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$ ; .002 mg/L of Arachidonic Acid ; 1.022 mg/L of Cholesterol; .070 mg/L of DL-alpha-Tocopherol-Acetate; 0.0520 mg/L of Linoleic Acid; 0.010 mg/L of Linolenic Acid; 0.010 mg/L of Myristic Acid; 0.010 mg/L of Oleic  
30 Acid; 0.010 mg/L of Palmitric Acid; 0.010 mg/L of Palmitic Acid; 100 mg/L of Pluronic F-68; 0.010 mg/L of Stearic Acid; 2.20 mg/L of Tween 80; 4551 mg/L of D-Glucose; 130.85 mg/ml of L- Alanine; 147.50 mg/ml of L-Arginine-HCL; 7.50 mg/ml of L-Asparagine- $\text{H}_2\text{O}$ ; 6.65 mg/ml of L-Aspartic Acid; 29.56 mg/ml of L-Cystine-2HCL- $\text{H}_2\text{O}$ ; 31.29 mg/ml of L-Cystine-2HCL; 7.35 mg/ml of L-Glutamic Acid; 365.0  
35 mg/ml of L-Glutamine; 18.75 mg/ml of Glycine; 52.48 mg/ml of L-Histidine-HCL- $\text{H}_2\text{O}$ ; 106.97 mg/ml of L-Isoleucine; 111.45 mg/ml of L-Leucine; 163.75 mg/ml of L-

Lysine HCL; 32.34 mg/ml of L-Methionine; 68.48 mg/ml of L-Phenylalanine; 40.0 mg/ml of L-Proline; 26.25 mg/ml of L-Serine; 101.05 mg/ml of L-Threonine; 19.22 mg/ml of L-Tryptophan; 91.79 mg/ml of L-Tyrosine-2Na-2H<sub>2</sub>O; 99.65 mg/ml of L-Valine; 0.0035 mg/L of Biotin; 3.24 mg/L of D-Ca Pantothenate; 11.78 mg/L of Choline Chloride; 4.65 mg/L of Folic Acid; 15.60 mg/L of i-Inositol; 3.02 mg/L of Niacinamide; 3.00 mg/L of Pyridoxal HCL; 0.031 mg/L of Pyridoxine HCL; 0.319 mg/L of Riboflavin; 3.17 mg/L of Thiamine HCL; 0.365 mg/L of Thymidine; and 0.680 mg/L of Vitamin B<sub>12</sub>; 25 mM of HEPES Buffer; 2.39 mg/L of Na Hypoxanthine; 0.105 mg/L of Lipoic Acid; 0.081 mg/L of Sodium Putrescine-2HCL; 55.0 mg/L of Sodium Pyruvate; 0.0067 mg/L of Sodium Selenite; 20uM of Ethanolamine; 0.122 mg/L of Ferric Citrate; 41.70 mg/L of Methyl-B-Cyclodextrin complexed with Linoleic Acid; 33.33 mg/L of Methyl-B-Cyclodextrin complexed with Oleic Acid; and 10 mg/L of Methyl-B-Cyclodextrin complexed with Retinal) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

### **Example 12: Construction of GAS Reporter Construct**

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class I, cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schindler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

| 5  | <u>Ligand</u>                    | <u>tyk2</u> | <u>JAKs</u> |             |             | <u>STATS</u> | <u>GAS(elements) or ISRE</u> |
|----|----------------------------------|-------------|-------------|-------------|-------------|--------------|------------------------------|
|    |                                  |             | <u>Jak1</u> | <u>Jak2</u> | <u>Jak3</u> |              |                              |
|    | <u>IFN family</u>                |             |             |             |             |              |                              |
|    | IFN-a/B                          | +           | +           | -           | -           | 1,2,3        | ISRE                         |
| 10 | IFN-g                            |             | +           | +           | -           | 1            | GAS (IRF1>Lys6>IFP)          |
|    | IL-10                            | +           | ?           | ?           | -           | 1,3          |                              |
|    | <u>gp130 family</u>              |             |             |             |             |              |                              |
| 15 | IL-6 (Pleiotrohic)               | +           | +           | +           | ?           | 1,3          | GAS (IRF1>Lys6>IFP)          |
|    | IL-11(Pleiotrohic)               | ?           | +           | ?           | ?           | 1,3          |                              |
|    | OnM(Pleiotrohic)                 | ?           | +           | +           | ?           | 1,3          |                              |
|    | LIF(Pleiotrohic)                 | ?           | +           | +           | ?           | 1,3          |                              |
|    | CNTF(Pleiotrohic)                | -/+         | +           | +           | ?           | 1,3          |                              |
| 20 | G-CSF(Pleiotrohic)               | ?           | +           | ?           | ?           | 1,3          |                              |
|    | IL-12(Pleiotrohic)               | +           | -           | +           | +           | 1,3          |                              |
|    | <u>g-C family</u>                |             |             |             |             |              |                              |
| 25 | IL-2 (lymphocytes)               | -           | +           | -           | +           | 1,3,5        | GAS                          |
|    | IL-4 (lymph/myeloid)             | -           | +           | -           | +           | 6            | GAS (IRF1 = IFP >> Ly6)(IgH) |
|    | IL-7 (lymphocytes)               | -           | +           | -           | +           | 5            | GAS                          |
|    | IL-9 (lymphocytes)               | -           | +           | -           | +           | 5            | GAS                          |
|    | IL-13 (lymphocyte)               | -           | +           | ?           | ?           | 6            | GAS                          |
| 30 | IL-15                            | ?           | +           | ?           | +           | 5            | GAS                          |
|    | <u>gp140 family</u>              |             |             |             |             |              |                              |
|    | IL-3 (myeloid)                   | -           | -           | +           | -           | 5            | GAS (IRF1>IFP>>Ly6)          |
| 35 | IL-5 (myeloid)                   | -           | -           | +           | -           | 5            | GAS                          |
|    | GM-CSF (myeloid)                 | -           | -           | +           | -           | 5            | GAS                          |
|    | <u>Growth hormone family</u>     |             |             |             |             |              |                              |
| 40 | GH                               | ?           | -           | +           | -           | 5            |                              |
|    | PRL                              | ?           | +/-         | +           | -           | 1,3,5        |                              |
|    | EPO                              | ?           | -           | +           | -           | 5            | GAS(B-CAS>IRF1=IFP>>Ly6)     |
| 45 | <u>Receptor Tyrosine Kinases</u> |             |             |             |             |              |                              |
|    | EGF                              | ?           | +           | +           | -           | 1,3          | GAS (IRF1)                   |
|    | PDGF                             | ?           | +           | +           | -           | 1,3          |                              |
|    | CSF-1                            | ?           | +           | +           | -           | 1,3          | GAS (not IRF1)               |

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994).), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5':GCGCCTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCG  
AAATGATTTCCTCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:3)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5':GCGGCAAGCTTTTGTCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5':CTCGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATG  
ATTTCCTCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC  
CTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGC  
CCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGC  
CTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTT  
TGCAAAAAGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

**Example 13: High-Throughput Screening Assay for T-cell Activity.**

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI + 10% serum with 1%Pen-Strep. Combine 2.5 mls of OPTI-MEM (Life Technologies)

with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

- During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final
- 5 concentration of  $10^7$  cells/ml. Then add 1ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentacin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

- 10 On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

- 15 Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100, 000 cells per well).

- After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12
- 20 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

- The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples
- 25 from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at -20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 4°C and serve as a source of material for repeating the assay on a specific well if desired.

- 30 As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

**Example 14: High-Throughput Screening Assay Identifying Myeloid Activity**

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells.

- 5 Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

- 10 To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et. al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^7$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

- 15 Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

- 20 Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

- 25 These cells are tested by harvesting  $1 \times 10^8$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

- 30 Add 50 ul of the supernatant prepared by the protocol described in Example 11. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

**Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCGAGGGATGACAGCGATAGAACCCCGG -3' (SEQ ID NO:6)

5' GCGAAGCTTCGCGACTCCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine

growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### **Example 16: High-Throughput Screening Assay for T-cell Activity**

NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF-  $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I-  $\kappa$ B is phosphorylated and degraded, causing NF-  $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF-  $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class 1 MHC.

Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating

diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40 early promoter sequence, and is flanked with an XhoI site:  
 5':GCGGCCTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGAC  
 TTTCCATCTGCCATCTCAATTAG:3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:  
 5':GCGGCAAGCTTTTGTCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2-. (Stratagene)

Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

5':CTCGAGGGGACTTTCCCGGGGACTTTCCGGGGACTTTCCGGGGACTTTCC  
 ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCCA  
 TCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT  
 AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCTCGGCCTCTGAGCTATTC  
 CAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGCTT:  
 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### **Example 17: Assay for SEAP Activity**

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### **Reaction Buffer Formulation:**

| # of plates | Rxn buffer diluent (ml) | CSPD (ml) |
|-------------|-------------------------|-----------|
| 10          | 60                      | 3         |
| 11          | 65                      | 3.25      |
| 12          | 70                      | 3.5       |
| 13          | 75                      | 3.75      |
| 14          | 80                      | 4         |
| 15          | 85                      | 4.25      |
| 16          | 90                      | 4.5       |
| 17          | 95                      | 4.75      |
| 18          | 100                     | 5         |
| 19          | 105                     | 5.25      |
| 20          | 110                     | 5.5       |
| 21          | 115                     | 5.75      |
| 22          | 120                     | 6         |

|    |     |       |
|----|-----|-------|
| 23 | 125 | 6.25  |
| 24 | 130 | 6.5   |
| 25 | 135 | 6.75  |
| 26 | 140 | 7     |
| 27 | 145 | 7.25  |
| 28 | 150 | 7.5   |
| 29 | 155 | 7.75  |
| 30 | 160 | 8     |
| 31 | 165 | 8.25  |
| 32 | 170 | 8.5   |
| 33 | 175 | 8.75  |
| 34 | 180 | 9     |
| 35 | 185 | 9.25  |
| 36 | 190 | 9.5   |
| 37 | 195 | 9.75  |
| 38 | 200 | 10    |
| 39 | 205 | 10.25 |
| 40 | 210 | 10.5  |
| 41 | 215 | 10.75 |
| 42 | 220 | 11    |
| 43 | 225 | 11.25 |
| 44 | 230 | 11.5  |
| 45 | 235 | 11.75 |
| 46 | 240 | 12    |
| 47 | 245 | 12.25 |
| 48 | 250 | 12.5  |
| 49 | 255 | 12.75 |
| 50 | 260 | 13    |

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

5 Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a  
10 fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

15 For adherent cells, seed the cells at 10,000 -20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50  $\mu$ l of 12  $\mu$ g/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100  $\mu$ l of buffer.

5 For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2-5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4  $\mu$ l of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension. The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100  
10  $\mu$ l/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200  $\mu$ l, followed by an aspiration step to 100  $\mu$ l final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the  
15 following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50  $\mu$ l. Increased emission at 530 nm indicates an extracellular signaling event which has resulted in an increase in the intracellular Ca<sup>++</sup> concentration.

20

#### **Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity**

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase  
25 RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

30 Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members  
35 of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium. Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer ((20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na<sub>3</sub>VO<sub>4</sub>, 2 mM Na<sub>4</sub>P<sub>2</sub>O<sub>7</sub> and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 4°C. The plate is then placed in a vacuum transfer manifold and the extract filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 4°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a

biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 5        The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg<sub>2+</sub> (5mM ATP/50mM MgCl<sub>2</sub>), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl<sub>2</sub>, 5 mM MnCl<sub>2</sub>, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the
- 10 components gently and preincubate the reaction mix at 30°C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mm EDTA and place the reactions on ice.

- Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction
- 15 mixture to a microtiter plate (MTP) module and incubating at 37°C for 20 min. This allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as
- 20 above.

- Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of
- 25 tyrosine kinase activity.

#### **Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

- As a potential alternative and/or compliment to the assay of protein tyrosine
- 30 kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase,
- 35 Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other

phosphoserine, phosphotyrosine, or phosphothreonine molecule, can be detected by substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp, (RT). The plates are then  
5 rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C  
10 until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts  
15 filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and  
20 Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.  
25

#### **Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is be isolated. cDNA is then generated from  
30 these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

35 PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies).

The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

5 PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

10 Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenin deoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

15 Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, Cv. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated  
25 disease.

#### **Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

30 A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

35 For example, antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10.

The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbounded polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbounded conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

#### **Example 23: Formulating a Polypeptide**

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally,

intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., *Biopolymers* 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., *J. Biomed. Mater. Res.* 15:167-277 (1981), and R. Langer, *Chem. Tech.* 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D-(-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., *Proc. Natl. Acad. Sci. USA* 82:3688-3692 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA* 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrans; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

**Example 24: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

5 Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily  
10 dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

**Example 25: Method of Treating Increased Levels of the Polypeptide**

15 Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5,  
20 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

**Example 26: Method of Treatment Using Gene Therapy**

25 One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is  
30 turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

5. The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

## SEQUENCE LISTING

## (1) GENERAL INFORMATION:

- 5 (i) APPLICANT: Human Genome Sciences, Inc., et al.  
(ii) TITLE OF INVENTION: 32 Human Secreted Proteins  
(iii) NUMBER OF SEQUENCES: 120
- 10 (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: Human Genome Sciences, Inc.  
(B) STREET: 9410 Key West Avenue  
(C) CITY: Rockville  
(D) STATE: Maryland  
15 (E) COUNTRY: USA  
(F) ZIP: 20850
- 20 (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage  
(B) COMPUTER: HP Vectra 486/33  
(C) OPERATING SYSTEM: MSDOS version 6.2  
25 (D) SOFTWARE: ASCII Text
- (vi) CURRENT APPLICATION DATA:  
30 (A) APPLICATION NUMBER:  
(B) FILING DATE: May 27, 1998  
(C) CLASSIFICATION:
- 35 (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER:  
(B) FILING DATE:
- 40 (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: A. Anders Brookes  
(B) REGISTRATION NUMBER: 36,373  
45 (C) REFERENCE/DOCKET NUMBER: P2006PCT
- (vi) TELECOMMUNICATION INFORMATION:  
50 (A) TELEPHONE: (301) 309-8504  
(B) TELEFAX: (301) 309-8439

55 (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 733 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCCCAAATCT TCTGACAAA CTCACACATG CCCACCGTGC CCAGCACCTG 60  
AATTGAGGG TGCACCGTCA GTCTTCCTCT TCCCCCAA ACCCAAGGAC ACCCTCATGA 120  
TCTCCCGGAC TCCTGAGGTC ACATGCGTGG TGGTGGACGT AAGCCACGAA GACCCTGAGG 180  
TCAAGTTCAA CTGGTACGTG GACGGCGTGG AGGTGCATAA TGCCAAGACA AAGCCGCGGG 240  
AGGAGCAGTA CAACAGCACG TACCGTGTGG TCAGCGTCCT CACCGTCCTG CACCAGGACT 300  
GGCTGAATGG CAAGGAGTAC AAGTGCAAGG TCTCCAACAA AGCCCTCCCA ACCCCCATCG 360  
AGAAAACCAT CTCCAAGCC AAAGGGCAGC CCCGAGAACC ACAGGTGTAC ACCCTGCCCC 420  
CATCCCGGGA TGAGCTGACC AAGAACCAGG TCAGCCTGAC CTGCCTGGTC AAAGGCTTCT 480  
ATCCAAGCGA CATCGCCGTG GAGTGGGAGA GCAATGGGCA GCCGGAGAAC AACTACAAGA 540  
CCACGCCTCC CGTGCTGGAC TCCGACGGCT CCTTCTTCCT CTACAGCAAG CTCACCGTGG 600  
ACAAGAGCAG GTGGCAGCAG GGAACGTCT TCTCATGCTC CGTGATGCAT GAGGCTCTGC 660  
ACAACCACTA CACGCAGAAG AGCCTCTCCC TGTCTCCGGG TAAATGAGTG CGACGGCCGC 720  
GACTCTAGAG GAT 733

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser  
1 5

## (2) INFORMATION FOR SEQ ID NO: 3:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 86 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GCGCCTCGAG ATTTCCCCGA AATCTAGATT TCCCCGAAAT GATTTCCTCG AAATGATTTC 60

5 CCGGAAATAT CTGCCATCTC AATTAG 86

10 (2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 base pairs

(B) TYPE: nucleic acid

15 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

20 GCGGCAAGCT TTTTGCAAAG CCTAGGC 27

25 (2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 271 base pairs

(B) TYPE: nucleic acid

30 (C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

35 CTCGAGATTT CCGGAAATC TAGATTTCCT CGAAATGATT TCCCCGAAAT GATTTCCTCG 60

AAATATCTGC CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC 120

40 GCCCTAACT CCGCCAGTT CCGCCATTC TCGCCCAT GGCTGACTAA TTTTITTTAT 180

TTATGCAGAG GCGAGGCG CCTCGGCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT 240

TTTTGGAGGC CTAGGCTTTT GCAAAAAGCT T 271

45

(2) INFORMATION FOR SEQ ID NO: 6:

50 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

55 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GCGCTCGAGG GATGACAGCG ATAGAACCCC GG 32

60

## (2) INFORMATION FOR SEQ ID NO: 7:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
10 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

15 GCGAAGCTTC GCGACTCCCC GGATCCGCCT C 31

## (2) INFORMATION FOR SEQ ID NO: 8:

- 20 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 base pairs  
(B) TYPE: nucleic acid  
25 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

30 GGGGACTTTC CC 12

## (2) INFORMATION FOR SEQ ID NO: 9:

- 35 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 73 base pairs  
(B) TYPE: nucleic acid  
40 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

45 GCGGCCTCGA GGGGACTTTC CCGGGACTT TCCGGGACT TTCCGGGACT TTCCATCCTG 60  
CCATCTCAAT TAG 73

## (2) INFORMATION FOR SEQ ID NO: 10:

- 50 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 256 base pairs  
55 (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CTCGAGGGGA CTTTCCCGG GACTTTCCGG GGACTTTCCG GGACTTTCCA TCTGCCATCT 60  
 CAATTAGTCA GCAACCATAG TCCCGCCCCCT AACTCCGCCC ATCCCGCCCC TAACTCCGCC 120  
 5 CAGTTCCGCC CATTTCCGC CCCATGGCTG ACTAATTTTT TTTATTTATG CAGAGGCCGA 180  
 GGCCGCCTCG GCCTCTGAGC TATTCCAGAA GTAGTGAGGA GGCTTTTTTG GAGGCCTAGG 240  
 10 CTTTTCGAAA AAGCTT 256

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1169 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GGGGCGCAAA TAGGGTCAGT GGGCGCCTTG GCGKTGTTCG TTGCGGTACC AGGTCCGCGT 60  
 GAGGGGTTCG GGGGTTCTGG GCAGGCACAA TGGCGTCTCG AGCAGGCCCG CGAGCGGCCG 120  
 RCACCGACGC AGCGAGCTTT CAGCACCGGG AGCGCGTCGC CATGCACTAC CAGATGAGTG 180  
 30 TGACCCTCAA GTATGAAATC AAGAAGCTGA TCTACGTACA TCTGGTCATA TGGCTGCTGC 240  
 TGGTTGCTAA GATGAGCGTG GGACACCTGA GGCTCTGTGC ACATGATCAG GTGGCCATGC 300  
 CCTATCAGTG GGAATACCCG TATTGCTGA GCATTTTGCC CTCTCTCTTG GGCCTTCTCT 360  
 35 CCTTCCCCG CAACAACATT AGCTACCTGG TGCTCTCCAT GATCAGCATG GGACTCTTTT 420  
 CCATCGCTCC ACTCATTTAT GGCAGCATGG AGATGTTCCC TGCTGCACAG CCTTCTACCG 480  
 40 CCATGGCAAG GCCTACCGTT TCCTCTTTGG TTTTCTGCCC GTTCCATCA TGTACCTGGT 540  
 GTTGGTGTG GCAGTGCAAG TGCATGCCTG GCAGTTGTAC TACAGCAAGA AGCTCCTAGA 600  
 45 CTCTTGGTTC ACCAGCACAC AGGAGAAGAA GCATAAATGA AGCCTCTTTG GGGTGAAGCC 660  
 TGGACATCCC ATCGAATGAA AGGACACTAG TACAGCGGTT CCAAAATCCC TTCTGGTGAT 720  
 TTTAGCAGCT GTGATGTTGG TACCTGGTGC AGACCCAGGC CAAAGTTCTG GAAAGCTCCT 780  
 50 TTTGCCATCT GCTGAGGTGG CAAACTATA ATTTATTCCT GGTGGCTAG AACTGGGTGA 840  
 CCAACAGCTA TGAAACAAAT TTCAGCTGTT TGAAGTGA CTTTGAGGTT TTTCTTTAAG 900  
 AATGAGCTTC GTCCTTGCTT CTACTCGGTC ATTCTCCCCA TTTCCATCCA TTACCCCTTA 960  
 55 GCCATTGAGA CTAAAGGAAA TAGGGAATAA ATCAAATTAC TTCATCTCTA GGTACGGGT 1020  
 CAGGAAACAT TTGGGCAGCT GCTCCCTTGG CAGCTGTGGT CTCTCTGCA AAGCATTTTA 1080  
 60 ATTAATAACC TCAATAAAGA TGCCCTGCCC ACAAAAAAAA AAAAAAAA AATTCGGGG 1140

GGGGCCCGG NAACCAATTN GCCCCTANA

1169

5

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1310 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

AATTCGGCAC GAGGCAGCGT CGCGCGGCC AGTTCCTTT TCCGGTCGGC GTGGTCTTGC 60

GAGTGGAGTG TCCGCTGTGC CCGGGCCTGC ACCATGAGCG TCCCGGCCTT CATCGACATC 120

AGTGAAGAAG ATCAGGCTGC TGAGCTTCGT GCTTATCTGA AATCTAAAGG AGCTGAGATT 180

TCAGAAGAGA ACTCGGAAGG TGGACTTCAT GTTGATTAG CTCAAATTAT TGAAGCCTGT 240

GATGTGTGTC TGAAGGAGGA TGATAAGAT GTTGAAAGTG TGATGAACAG TGTGGTATCC 300

CTACTCTTGA TCCTGGAACC AGACAAGCAA GAAGCTTTGA TTGAAAGCCT ATGTGAAAAG 360

CTGGTCAAAT TTCGCGAAGG TGAACGCCCG TCTCTGAGAC TGCAGTTGTT AAGCAACCTT 420

TTCCACGGGA TGGATAAGAA TACTCCTGTA AGATACACAG TGTATTGCAG CCTATTATAA 480

GTGGCAGCAT CTGTGGGGC CATCCAGTAC ATCCCAACTG AGCTGGATCA AGTTAGAAAA 540

TGGATTCTTG ACTGGAATCT CACCACTGAA AAAAAGCACA CCCTTTTAAG ACTACTTTAT 600

GAGGCACCTG TGGATTGTAA GAAGAGTGAT GCTGCTTCAA AAGTCATGGT GGAATTGCTC 660

GGAAGTTACA CAGAGGACAA TGCTTCCCAG GCTCGAGTTG ATGCCACAG GTGTATTGTA 720

CGAGCATTGA AAGATCCAAA TGCATTTCTT TTTGACCACC TTCTTACTTT AAAACCAGTC 780

AAGTTTTTGG AAGGCGAGCT TATTCATGAT CTTTAAACCA TTTTGTGAG TGCTAAATTG 840

GCATCATATG TCAAGTTTTA TCAGAATAAT AAAGACTTCA TTGATTCACT TGGCCTGTTA 900

CATGAACAGA ATATGGCAA AATGAGACTA CTTACTTTTA TGGGAATGGC AGTAGAAAAT 960

AAGGAAATTT CTTTGTACAC AATGCAGCAA GAACTTCAGA TTGGAGCTGA TGATGTTGAA 1020

GCATTGTGTA TTGACGCCGT AAGAACTAAA ATGGTCTACT GCAAAATTGA TCAGACCCAG 1080

AGAAAAGTAG TTGTCAGTCA TAGCACACAT CGGACATTTG GAAAACAGCA GTGGCAACAA 1140

CTGTATGACA CACTTAATGC CTGGAAACAA AATCTGAACA AAGTGAAAAA CAGCCTTTTG 1200

AGTCTTTCTG ATACCTGAGT TTTTATGCTT ATAATTTTGT TTCTTTGAAA AAAAAGCCCT 1260

AAATCATAGT AAAACATTAT AACTAAAAA AAAAAAAAAA AAAAAAAAAA 1310

60

## (2) INFORMATION FOR SEQ ID NO: 13:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1139 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

|   |      |
|---|------|
| AGGGCANACT TACAGAGATA TCATATGAGA TCACCCCTCG CATTCGTGTC TGGCGCCAGA | 60   |
| CCCTCGAGCG GTGCCGAGC GCASCCAGGT GTGCTTGTGC CTGGGCCAGC TGGAGAGGTC  | 120  |
| CATTGCCTGG GANGAAGTCT GTCAACAAAG TGACATGTCT AGTCTGCCG AAGGGTGACA  | 180  |
| ATGATGAGTT TCTTCTGCTT TGTGATGGT GTRACCGTGG CTGCCACATT TACTGCCATC  | 240  |
| GTCCCAAGAT GGAGGCTGTC CCAGAAGGAG ATTGGTCTG TACTGTCTGT TTGGCTCAGC  | 300  |
| AGGTGGAGGG AGAATTCAC TACAAGCCTG GTTTCCTCAA GCGTGGCCAG AAGCGGAAAA  | 360  |
| GTGGTTATTC GCTGAACCTC TCAGAGGGTG ATGGCGCCG ACGCCGGTA CTGTTGAGGG   | 420  |
| GCCGAGAAAG CCCAGCAGCA GGGCCTCGGT ACTCGGAAGA AGGGCTCTCC CCCTCCAAGC | 480  |
| GGCGGCGACT CTCTATGCGG AACCACCACA GTGATCTCAC ATTTTGCGAG ATTATCCTGA | 540  |
| TGGAGATGGA GTCCCATGAT GCAGCCTGGC CTTTCCTAGA GCCTGTGAAC CCACGTTTGG | 600  |
| TGAGTGGGTA CCGGCGCATC ATCAAAAATC CTATGGATT TCCACCATG CGGGAGCGGC   | 660  |
| TGCTCAGGG AGGGTACACC AGCTCAGAG AGTTTGCGG TGATGCCCTC CTGGTATTTG    | 720  |
| ACAACCTCCA GACTTTCAAC GAGGATGACT CTGAAGTAGG CAAGGCTGGG CACATCATGC | 780  |
| GCCGCTTCTT CGAGAGCCGC TGGGAGGAGT TTTATCAGG AAAACAGGCC AATCTGTGAG  | 840  |
| GCAAGGGAGG TGGGGAGTCA CTTGTGGCA TCTCCCCCA CTTCCAAAC AAAAACCTGC    | 900  |
| CATTTTCACC TGCTGATGCT GCCCTGGGTC CAGACTCAAG TCAGATACAA CCCTGATTTT | 960  |
| TGACCTTACC CTTGGCAGT GCCACATCC TCTTATTCCT ACATCCCTTT CTCCCTTCCC   | 1020 |
| TCCTCTTGCT CCTCAAGTAA GAGGTGCAGA GATGAGGTCC TTCTGGACTA AAAGCCAAAA | 1080 |
| AAAGAAAGAA AAAAATAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA | 1139 |

## (2) INFORMATION FOR SEQ ID NO: 14:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2271 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

|    |   |      |
|----|---|------|
| 5  | GTTCGGGGG ATGCCAGCTC ACTTCTCGGA CAGCGCCCAG ACTGAGGCCT GCTACCACAT  | 60   |
|    | GCTGAGCCGG CCCAGCCGC CACCGACCC CTCCTGCTC CAGCGTCTGC CACGCCCAG     | 120  |
| 10 | CTCCCTGTCA GACAAGACCC AGCTCCACAG CAGGTGGCTG GACTCGTCGC GGTGTCTCAT | 180  |
|    | GCAGCAGGGC ATCAAGCCG GGGACGCACT CTGGCTGCGC TTCAAGTACT ACAGCTTCTT  | 240  |
|    | CGATTGATGAT CCCAAGACAG ACCCGTGGC GCTGACACAG CTGTATGAGC AGGCCCGGTG | 300  |
| 15 | GGACCTGCTG CTGGAGGAGA TTGACTGCAC CGAGGAGGAG ATGATGGTGT TTGCCGCCCT | 360  |
|    | GCAGTACCAC ATCAACAAGC TGTCCAGAG CGGGGAGGTG GGGGAGCCG CTGGCACAGA   | 420  |
| 20 | CCCAGGGCTG GACGACCTGG ATGTGGCCCT GAGCAACCTG GAGGTGAAGC TGGAGGGGTC | 480  |
|    | GGCGCCCA GATGTGCTGG ACAGCCTCAC CACCATCCCA GAGCTCAAGG ACCATCTCCG   | 540  |
|    | AATCTTTCGG CCCCGGAAGC TGACCCTGAA GGGCTACCGC CAACACTGGG TGGTGTCAA  | 600  |
| 25 | GGAGACCACA CTGTCTACT ACAAGAGCCA GGACGAGGCC CTGGGGACC CCATTTCAGCA  | 660  |
|    | GCTCAACCTC AAGGGCTGTG AGGTGGTTCC CGATGTTAAC GTCTCCGGCC AGAAGTCTG  | 720  |
| 30 | CATTAACTC CTAGTGCCCT CCCCTGAGGG CATGAGTGAG ATCTACCTGC GGTGCCAGGA  | 780  |
|    | TGAGCAGCAG TATGCCGCT GGATGGCTGG CTGCCGCTG GCCTCCAAAG GCCGCACCAT   | 840  |
|    | GGCCGACAGC AGCTACACCA GCGAGGTGCA GGCCATCCTG GCCTTCCTCA GCCTGCAGCG | 900  |
| 35 | CACGGGCACT GGGGGCCCG GCAACCACCC CCACGGCCCT GATGCCTCTG CCGAGGGCCT  | 960  |
|    | CAACCCCTAC GGCTCGTTG CCCCCGTTT CCAGCGAAG TTCAAGGCCA AGCAGCTCAC    | 1020 |
| 40 | CCCACGGATC CTGGAAGCCC ACCAGAATGT GGCCAGTTG TCGCTGGCAG AGGCCCAGCT  | 1080 |
|    | GCGCTTCATC CAGGCCTGGC AGTCCCTGCC CGACTTCGGC ATCTCCTATG TCATGGTCAG | 1140 |
|    | GTTCAGGGC AGCAGGAAAG ACGAGATCCT GGGCATCGCC AACAACCGAC TGATCCGCAT  | 1200 |
| 45 | CGACTTGGCC GTGGGCGAGC TGGTCAAGAC CTGGCGTTTC AGCAACATGC GCCAGTGGAA | 1260 |
|    | TGTCAACTGG GACATCCGGC AGGTGGCCAT CGAGTTTGAT GAACACATCA ATGTGGCCTT | 1320 |
| 50 | CAGCTGCGTG TCTGCCAGCT GCGAATGTGT ACACGAGTAT ATCGGGGGCT ACATTTCCT  | 1380 |
|    | GTGACGCGG GAGCGGGCCC GTGGGAGGA GCTGGATGAA GACCTCTTCC TGCAGCTCAC   | 1440 |
|    | CGGGGGCCAT GAGGCCTTCT GAGGGCTGTC TGATTGCCCC TGCCCTGCTC ACCACCTGT  | 1500 |
| 55 | CACAGCCACT CCCAAGCCCA CCCCCACAGG GGCTCACTGC CCCACACCG CTCCAGGCAG  | 1560 |
|    | GCACCCAGCT GGGCATTTC CTTGCTGTCA CTGACTTTGT GCAGGCCAAG GACCTGGCAG  | 1620 |
| 60 | GGCAGACGC TGTACCATCA CCCAGGCCAG GGATGGGGT GGGGTCCCT GAGCTCATGT    | 1680 |

GGTGCCCCCT TTCCTGTCT GAGTGGCTGA GGCTGATACC CCTGACCTAT CTGCAGTCCC 1740  
 CCAGCACACA AGGAAGACCA GATGTAGCTA CAGGATGATG AAACATGGTT TCAAACGAGT 1800  
 5 TCTTTCTTGT TACTTTTAA AATTTCCTTT TTATAAATTA ATATTTTATT GTTGGATCCT 1860  
 CCTCCTTTCT CTGGAGCTGT GCTTGGGGCT ACTCTGACAC TCTGTCTCTT CATCACCAGC 1920  
 10 CAAGGAAAGG GGCTTTCCTG ATAAAGACAA GAGTTGGTTA GAGAAAGGA CACCTAAGTC 1980  
 AGTCTAGGCT TGAAGCTAG GAGAGAGGTG AGGGCAGAAG GGCACAGCTT TCAGGAACAA 2040  
 GGAATAGGGG CTGGGGTKGT KGTTCACAG GGTAGGCGTA CCTGCAGGGC CTCCTTGAAG 2100  
 15 TACTTGGGAA GGAGGAAGCC ATCAGTATTC CCTGGAGTCA GAATCACCCC ATTGGCAGAG 2160  
 CGGAAGAAGG GTATTCCATC TGCTGACAGA GCCAGAGATG TGAATCATGC CCTCCCGAA 2220  
 20 GGCAAAGTCA GCTCCTGCTT TGTCCAGACT CACCTGCCAG AGCCAGGGGT C 2271

25 (2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 626 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

35 ACAACAAACA TCGAAATCG ANTATGTGCC CCGAAAAGTC GGAACGCAGG CAATCAGTCC 60  
 GCACGMGCGC AAGTTCAACA TGAAGATGAT ATGAGGCCGG GCGGGGGGGC AGGGACCCCC 120  
 GGGCGGCCGG GCAGGGGAAG GGGCCTGGCC GCCACCTGCT CACTCTCCAG TCCTTCCAC 180  
 40 CTCCTCCCTA CCTTCTACA CACGTTCTCT TTCTCCCTCC CGCCTCCGTC CCCTGCTGCC 240  
 CCCC GCCAGC CCTCACCACC TGCCCTCCTT CTACCAGGAC CTCAGAAGCC CAGACCTGGG 300  
 GACCCACCT ACACAGGGGC ATTGACAGAC TGGAGTTGAA AGCCGACGAA CCGACACGCG 360  
 45 GCAGAGTCAA TAATCAATA AAAAAGTTAC GAACTTTCTC TGTAACCTGG GTTCAATAA 420  
 TTATGGATTT TTATGAAAAC TTGAAATAAT AAAAAGAGAA AAAAATATT TCCTATAGCT 480  
 50 AGTCGGAATG CAACTTTTG ACGTCTGAT TGCTCCAGGG CCCTCTTTCC AACTCAGTTT 540  
 CTGTGTTTTT CTCTCTCTCC TCCTCCTCTT CTCTCTCTT TCTTCTCTT NCCCCATGGG 600  
 55 GGAGGGGTTT ATTCAGGGAA AACAGG 626

60 (2) INFORMATION FOR SEQ ID NO: 16:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2118 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

|    |  |      |
|----|--|------|
| 10 | TTTTCCAGCC ATGTCATAA TTGTGAATC CTACCAACTA TTGACAGAAT ACAGAGTTGA    | 60   |
|    | TTTTTTAATA AAAAGTTATA TATAATTATC CCTTTAATTA AAGGGAGCAA AGGGGCGTTC  | 120  |
|    | CACATGGACA GAGGCTTGA CCGAGGCCTG GTCACAGCAG CGAGCATCCA GGGTTTGCAG   | 180  |
| 15 | GGACGATGTT ACAGACTCTG TTTCTGCCT GCGGTTTCAC TTGTGTCTGC TCCTAGCCTG   | 240  |
|    | TGCTCTGCCA GCAGCACAGA CATCTGCTCC ATCAGACCTC TTCCATTTTG CACAGGGAGT  | 300  |
|    | GCAGGAGGTG AATGTTCACT TTCTGTCTC CAGTGTCACT GTTCTGTTTC CACGGGATGG   | 360  |
| 20 | AAAGCGCATG GGCCTGTGTC CATGTAGAT TTCCTTCTAG ATTCTGTGT ACACACACTT    | 420  |
|    | GATTGTCTG GATGAATGTC TTTTTAATA CTCGAAAAT TTCATCATCT AAGAAAATGA     | 480  |
| 25 | TTCCATACAA ATAATCAGC ACACAAGTGA CCCAGGACAT ATGCCTGCCA AAGGGATCTG   | 540  |
|    | TTAGAAGGCT GCCTTCTCAT GCGCATGTC ACTTGGATCT TGTGGTGAGG ACGGCCCCAT   | 600  |
|    | CTTTCTTGCC ACAGATTGAG GCCACTTTTG AGCAAGGGAG ATCCTGGAGT TAAGACAGGT  | 660  |
| 30 | GTGGGGGCA GCCTGTATTT TACCCTAGGG GCAGGTCTGC ATGGTGACCC CACATYGCAC   | 720  |
|    | TGGTAAACCA TTTAGTCCC ACTCTTCATC CTGGAAGTGG GAACTGGAGT CCCACCCACA   | 780  |
| 35 | GTGCATTCAG AAAGCATGCT GTGTGGGGC TGCTTCTCAG GAGGCCAGGC CCTTCTGAGC   | 840  |
|    | GGAACCGTCC TGGAGAGAGC CTGCCCTCGT TTCCAGGCTG CAGCCGTAAC GCACTTTCTC  | 900  |
|    | CCAGGCTGAG GCGGGGTGTT CTGGGGTCTC TGCCCTCTGT CGGCCCTGCT TCCTGCCAGG  | 960  |
| 40 | ACGTGGCCTC TTCCGATCCT TTTCTCTCAG AACTGGAGG TCTCTTCTGC CATTGTGCTG   | 1020 |
|    | GTCCCATCCC AAGAAATGTA GGACAGAGAC CAACTGGGT CGCGGACAC AAAGTCCATC    | 1080 |
| 45 | CAGGACCCAG GCCGAGAGG GAGCAGGAAG AGATGCTGAT AGTTTGATCT AGAAACCAGC   | 1140 |
|    | AGCTACTGGC TCAAATTCAG GTTCTGGCGT CAAATAGCGA CATTTCCAGT TTCTCTTAAA  | 1200 |
|    | AACCGTGTTC GTTTTCAGTT GGGATAGGCT TGTTTTGTCT GTTGAAAATG TTTCTAGTTT  | 1260 |
| 50 | TTTTTCTTTC ATTTTCTCT CATTCATTT CTGCCTTAAC TTTAGTTTGT TCACAGGGAG    | 1320 |
|    | GCAAAGCTGA CATGAACCTT TTGTCGTGGG ACTTCAGGCC ACATTGGCTT GAAGGCATTTC | 1380 |
| 55 | GTTCCTTCT GGGGTGGGA CAGGCCCTCA TGGCAGGCTT GTTCCCGTGG CTCTGAGCGA    | 1440 |
|    | GGCCTCTTCC TGCTGGGCTC CCAGACTCCT GCATCCAGGC CCCACCTTC TCGGCTTCTG   | 1500 |
| 60 | GTTTTCTTTT CTTTTTGTA GAACACAACA TCTACCATTC AGTTAAACCT TCTTTATCTC   | 1560 |

CTCCTYTGGC ATCCATTTTT CCAAAGAAGA GTCGAGTCCT CTGAGGTCTG TGCTTGAAAR 1620  
 CCGTCCGAAG GCATTCTTGT TAGCTTTGCT TTTCTCCCA TATCCCAAGG CGAAGCGCTG 1680  
 5 AGATTCTTCC ATCTAAAAA CCCTCGACCC GAAACCTCA CCAGATAAAC TACAGTTTGT 1740  
 TTAGGAGGCC CTGACCTTCA TGGTGTCTTT GAAGCCCAAC CACTCGGTTT CCTTCGGATT 1800  
 10 TTCCTCCCTT TGTTCGGGGT TTGGTTTGGC TCCTCTGTGT GTGTCCGTAT CTGTTCGGT 1860  
 GTCCTCGAGG TTGAGCTTCA CTCCACTGCG GCAGAGGCAG CGTGCACT CGGATTGCT 1920  
 ACGTTTCTAT ATATCTTGAA GCTAAATGTA TATATGAGTA GTTTGCCATG AGATAACACA 1980  
 15 GTGTAAACAG TAGACACCA GAAATCGTGA CTTCTGTGTT CTCTCCATT T GAGTATTTTG 2040  
 TAATTTTTT GAAATATTG TGGACATAA TAAACCAAG CTACACTACA AAAAAAAAAA 2100  
 20 AAAAAAAGT GAGACTAG 2118

25 (2) INFORMATION FOR SEQ ID NO: 17:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1076 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 30 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

35 GCCCAAGGAG CTCAGCTTCG CCCGCATCAA GGCCTTGAG TGCCTGGAGA GCACCGGGCG 60  
 CCACATCTAC TTCACGCTGG TGACCGAAGG GWGCGGCGAG ATCGACTTCC GCTGCCCCCT 120  
 GGAAGATCCC GGCTGGAACG CCCAGATCAC CCTAGGCCTG GTCAAGTTCA AGAACCAGCA 180  
 40 GGCCATCCAG ACAGTGGGG CCCGGCAGAG CCTCGGGACC GGGACCCCTG TGTCTAAAC 240  
 CACCGGGGCG ACCATCTTTC CTTCATGCTA CCCACCACCT CAGTGTGAG GTCAAGGCAG 300  
 45 CTTCTGTGTT CCCTCTGGCT TGTGGGGCA CGGCTGTSYT CCATGTGGCA AGGTGGAAGG 360  
 CATGGACGTG TGGAGGAGGC GCTGGAGCTG AAGGAATGGA CGAGCCCTGG GAGGAGGGCA 420  
 GAAGGCTACG CAGGCTGAG GATGAAGATG CAGCCCCCTG ATGGTCCAG ACTCTCAGGA 480  
 50 CATGCCCAGC TCAGGGGCTT CGAGCCACAG GCCTGGCCTC ATATGGCATG AGGGGAGCT 540  
 GGCATAGGAG CCCCTCCCT GCTGTGGTCC TGCCCTCTGT CCTGCAGACT GCTCTTAGCC 600  
 55 CCCTGGCTTT GTGCCAGGC TGGAGAGGG CAGTCCCCCA TGGGTGCGG AGCCAACGCC 660  
 TCAGGAATCA GGAGCCAGC CTGGTACCAA AAGGAGTACC CAGGGCCTGG TACCCAGGCC 720  
 CACTCCAGAA TGGCCTCTGG ACTCACCTTG AGAAGGGGA GCTGCTGGGC CTAAAGCCCA 780  
 60 CTCCTGGGGG TCTCTGCTG CTTAGGTCCT TTTGGGACCC CCACCCATCC AGGCCCTTTC 840

5 TTTGCACACT TCTTCCCCCA CCTCTAYGCA TCTTCCCCCC ACTGCGGTGT TCGGCCTGAA 900  
 GGTGGTGGGG GTGAGGGGGG GTTGGCCAT TAGCATTTCA TGTCTTTCCC CAAATGAAGA 960  
 TGCCCTGCAA AGGGCAGTNA ACCACAAAAA AAAAAAAAAA AAAACNTGG GGGGGGGGCC 1020  
 CCGTTAACCA TTTTGGCCTN ATAGGGGGGN GGTTTTAA AATTAATGG GCCCGG 1076

10

(2) INFORMATION FOR SEQ ID NO: 18:

15

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1379 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

GGCACGAGCA CCTCCACCA CCTCCCTGAA CTTCATCTG ATCGACTTCA ACTTGCTGAT 60  
 25 GGTGACCACC ATCGTTCTGG GCCGCCGCTT CATTTGGTCC ATCGTGAAGG AGGCCTCTCA 120  
 GAGGGGAAG GTCTCCCTCT TTCGCTCCAT CCTGCTGTC CTCACTCGCT TCACCGTTCT 180  
 30 CACGGCAACA GGCTGGAGTC TGTGCCGATC CCTCATCCAC CTCTTCAGGA CCTACTCCTT 240  
 CCTGAACCTC CTGTTCTCTT GCTATCCGTT TGGGATGTAC ATTCCGTTCC TGCARCTGAA 300  
 TTKGAMCTY CGSAAGACAA GCCTCTTCAA CCACATGGCC TCCATGGGGC CCCGGGAGGC 360  
 35 GGTGAGTGGC CTGGCAAAGA GCGGGGACTA CCTCCTGACA CTGCGGGAGA CGTGGAAGCA 420  
 GCACASAAGA CAGCTGTATG GCCCGGACGC CATGCCACC CATGCCTGCT GCCTGTGCGC 480  
 40 CAGCCTCATC CGCAGTGAGG TGGAGTTCTT CAAGATGGAC TTCAACTGGC GCATGAAGGA 540  
 AGTGCTCGTS AGCTCCATGC TGAGCGCCTA CTATGTGGCC TTTGTGCCTG TTTGGTTCGT 600  
 GAAGAACACA CATTACTATG ACAAGCGCTG GTCCTGTGNA ACTCTTCCTG CTGGTGTCCA 660  
 45 TCAGCACCTC CGTGATCTC ATGCAGCACC TGCTGCNTGC CAGCTACTGT GACCTGCTGC 720  
 ACAAGGCCGC CGCCCATCTG GGCTGTTGGC AGAAGGTGGA CCCAGCGCTG TGCTCCAACG 780  
 TGCTGCAGCA CCCGTGGACT GAAGAATGCA TGTGGCCGCA GGGCGTGCTG GTGAAGCACA 840  
 50 GCAAGAACGT CTACAAAGCC GTAGGCCAMW ACAAMGTGGC TATCCCTCTT GACGTCTCCC 900  
 ACTTCCGCTT CCAKTTCTTT TTCAGCAAAC CCCTGCGGAT CCTCAACATC CTCCTGCTGC 960  
 55 TGGAGGGGCG TGTCATTGTC TATCAGCTGT ACTCCCTAAT GTCTCTGAA AAGTGGCACC 1020  
 AGACCATCTC GCTGGCCCTC ATCCTCTTCA GCAACTACTA TGCTTCTTC AAGCTGCTCC 1080  
 GGGACCGCTT GGTATTGGGC AAGGCCTACT CATACTCTGC TAGCCCCCAG AGAGACCTGG 1140

60

|    |   |      |
|----|---|------|
|    | ACCACCGTTT CTCCTGAGCC CTGGGGTCAC CTCAGGGACA GCGTCCAGGC TTCAGCAAGG | 1200 |
|    | GCTCCCTGGC AAGGGGCTGT TGGGTAGAAG TGGTGGTGGG GGGGACAAAA GACAAAAAAA | 1260 |
| 5  | TCCACCAGAG CTTTGTATTT TTGTTACGTA CTGTTTCTTT GATAATTGAT GTGATAAGGA | 1320 |
|    | AAAAAGTCCT ATTTTATATAC TCCCAANMAA AAAAAAAAAA NAAAAAGCGG CCGAAAGCT | 1379 |
| 10 |   |      |
|    | (2) INFORMATION FOR SEQ ID NO: 19:                                |      |
| 15 | (i) SEQUENCE CHARACTERISTICS:                                     |      |
|    | (A) LENGTH: 1337 base pairs                                       |      |
|    | (B) TYPE: nucleic acid  |      |
|    | (C) STRANDEDNESS: double  |      |
|    | (D) TOPOLOGY: linear  |      |
| 20 | (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:                         |      |
|    | CTGGTGTGG GCCTGAGCCN CCTCAACAAC TCCTACAAC TCACTTTCCA CGTGGTGATC   | 60   |
| 25 | GGCTCTCAGG CGGAAGAAGG CCAGTACAGC CTGAACCTCC ACAACTGCAA CAATTCAGTG | 120  |
|    | CCAGGAAAGG AGCATCCATT CGACATCAGC GTGATGATCC GGGAGAAGAA CCCGATGGC  | 180  |
|    | TTCTGTGCGG CAGCGGAGAT GCCCTTTTC AAGCTCTACA TGGTCATGTC CGCTGCTTC   | 240  |
| 30 | CTGGCCGCTG GCATCTCTG GGTGTCCATC CTCTGCAGGA ACACGTACAG CGTCTTCAAG  | 300  |
|    | ATCCACTGGC TCATGGCGGC CTGGCCCTTC ACCAAGAGCA TCTCTCTCCT CTTCCACAGC | 360  |
| 35 | ATCAACTACT ACTTCATCAA CAGCCAGGGG CCACCCCATC GAAGGCCTTG CCGKCATGTA | 420  |
|    | CTACATCGCA CACCTGCTGA AGGGCGCCCT CCTCTTCATC ACCATCGCCC TGATTGGCTC | 480  |
|    | AGGCTGGGCT TCATCAAGTA CGTCTGTG GATAAGGAGA AGAAGGTCTT TGGGATCGTG   | 540  |
| 40 | ATCCCCATGC AGGTCTGGC CAACGTGGCC TACATCATCA TCGAGTCCCG CGAGGAAGGC  | 600  |
|    | GCCACGAACT ACGTGTGTG GAAGGAGATT TTGTTCTGG TGGACCTCAT CTGCTGTGGT   | 660  |
| 45 | GCCATCCTGT TCCCCGTAGT CTGGTCCATC CGGCATCTCC AGGATGCGTC TGGCACAGAC | 720  |
|    | GGGAAGGTGG CAGTGAACCT GGCCAAGCTG AAGCTGTTCC GGCATTACTA TGTGATGGTC | 780  |
|    | ATCTGCTACG TCTACTTCAC CCGCATCATC GCCATCCTGC TGCAGGTGGC TGTGCCCTTT | 840  |
| 50 | CAGTGGCAGT GGCTGTACMA GCTCTTGGTG GARGGCTCCA CCCTGGCCTT CTTGCTGCTC | 900  |
|    | ACGGGCTACA AGTTCAGCC CACAGGGAAC AACCGTACC TGCAGCTGCC CCAGGAGGAC   | 960  |
| 55 | GAGGAGGATG TTCAGATGGA GCAAGTAATG ACGGACTCTG GGTTCGGGA AGGCCTCTCC  | 1020 |
|    | AAAGTCAACA AAACAGCCAG CGGGCGGGAA CTGTTATGAT CACCTCCACA TCTCAGACCA | 1080 |
|    | AAGGTCGTC CTCCCCAGC ATTTCTCACT CCTGCCCTTC TTCCACAGCG TATGTGGGGA   | 1140 |
| 60 | GGTGGAGGGG TCCATGTGGA CCAGGCGCCC AGCTCCCGGG ACSCCGGTT CCGGACAAGC  | 1200 |

CCATTTGGAA GAAGAGTCCC TTCCTCCCC CAAATATTGG GCAGCCCTGT CCTTACCCCG 1260  
 5 GGACCACCCC TCCCTTCCAG CTATGTGTAC AATAATGACC AATCTGTTTG GCTAAAAAAA 1320  
 AAAAAAAAAA AACTCGA 1337

10

(2) INFORMATION FOR SEQ ID NO: 20:

(i) SEQUENCE CHARACTERISTICS:

15

- (A) LENGTH: 1390 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

GCGGTTTGG TTCCCGGTTG GTGCTTCCTG TTCGCAGCTG CGGCACCTCA AGGTTACTGA 60  
 CTMTTATGA TGTGTGGTGG CTATGAGACT ATAGAWGCRS RSGRRGATGA TYTTTATCGA 120  
 25 GATGAGTCAT CTAGTGAAC TGTGTGTGAT AGTGAGGTGG AATTTCAACT CTATAGCCAA 180  
 ATTCATTATG CCCAAGATCT TGATGATGTC ATCAGGGAGG AAGAGCATGA AGAAAAGAAC 240  
 TCTGGGAATT CGGAATCTTC GAGTAGTAAA CCAAATCAGA AGAAGCTAAT CGTCCTTTCA 300  
 30 GATAGTGAGG TCATCCAGCT GTCAGATGGG TCAGAGGTCA TCACTTTGTC TGATGAAGAC 360  
 AGTATTTATA GATGTAAAGG AAAGAATGTT AGAGTTCAAG CACAAGAAAA TGCCCATGGT 420  
 35 CTMTCTCTT CTCTTCAATC TAATGAGCTG GTTGATAAGA AATGCAAGAG TGATATTGAG 480  
 AAGCCTAAAT CTGAAGAGAG ATCAGGTGTA ATCCGAGAGG TCATGATTAT AGAGTCACT 540  
 TCAAGTGAAG AGGAAGAGAG CACCATTTCA GAAGGTGATA ATGTGGAAAG CTGGATGCTA 600  
 40 CTGGGATGTG AAGTAGATGA TAAAGATGAT GATATCCTTC TCAACCTTGT GGGATGTGAA 660  
 AACTCTGTTA CTGAAGGAGA AGATGGTATA AACTGGTCCA TCAGTGACAA AGACATTGAG 720  
 45 GCCCAGATAG CTAATAACCG AACACCTGGA AGATGGACCC AGCGGTACTA TTCAGCCAAC 780  
 AAAACATTA TCTGTAGAAA TTGTGACAAA CGTGGTCATT TATCAAAAAA CTGCCCCCTA 840  
 CCACGAAAAG TTCGTCGCTG CTTCTGTGTC TCCAGGAGAG GACATCTCCT GTATTCTGT 900  
 50 CCAGCCCCC TTTGCGAATA CTGTCCTGTG CTAAGATGT TGGACCACTC ATGTCTTTTC 960  
 AGACATTCCT GGGATAAACA GTGTGACCGA TGTCATATGC TAGGCCACTA TACAGATGCT 1020  
 55 TGCACAGAAA TCTGGAGGCA GTATCACCTA ACGACCAAAC CTGGACCACC CAAAAGCCG 1080  
 AAGACCCCTT CAAGACCATC AGCCTTAGCA TATTGCTATC ACTGCGCGCA AAAAGGCCAT 1140  
 60 TATGGACACG AATGTCCAGA AAGAGAAGTG TATGACCCGT CTCCAGTATC TCCATTTCATC 1200

TGCTACTATG RTGACAAATA TGAAATTCAG GAGAGAGAAA AGAGACTAAA ACAAAAAATA 1260  
 AAAGTANICA AGAAAAATGG GGTATATCCCA GAGCCATCCA AGCTACCTTA TATAAAAGCA 1320  
 5 GCAAATGAGA ACCCCCACCA TGATATAAGG AAGGGCCGTG CCTCATGGAA AAGCAACAGG 1380  
 TGGCCTCAAG 1390

10

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:  
 15 (A) LENGTH: 1431 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

GCCTGCAGTC GACACTAGTG GATCCAAAGA ATTCCGCCCTG TGCGAGTAGG CGCTTGGGCA 60  
 CTCAGTCTCC CTGGCGAGCG ACGGGCAGAA ATCTCGAACC AGTGGAGCGC ACTCGTAACC 120  
 25 TGGATCCAG AAGGTCGCGA AGGCAGTACC GTTTCCTCAG CGGCGGACTG CTGCAGTAAG 180  
 AATGTCITTTT CCACCTCATT TGAATCGCCC TCCCATGGGA ATCCCAGCAC TCCCACCAGG 240  
 30 GATCCCACCC CCGCAGTTTC CAGGATTTCC TCCACCTGTA CCTCCAGGGA CCCCAATGAT 300  
 TCCTGTACCA ATGAGCATTA TGGCTCCTGC TCCAACCTGC TTAGTACCCA CTGTGTCTAT 360  
 GGTGGAAG CATTTGGGCG CAAGAAAGGA TCATCCAGGC TTAAAGGCTA AAGAAAATGA 420  
 35 TGAAAATTGT GGTCTACTA CCACTGTITTT TGTGGAAC ATTTCCGAGA AAGCTTCAGA 480  
 CATGCTTATA AGACAACCTT TAGCTAAATG TGGTTTGGTT TTGAGCTGGA AGAGAGTACA 540  
 40 AGGTGCTTCC GGAAAGCTTC AAGCCTTCGG ATICTGTGAG TACAAGGAGC CAGAATCTAC 600  
 CCTCCGTGCA CTCAGATTAT TACATGACCT GCAAATTGGA GAGAAAAAGC TACTCGTTAA 660  
 AGTTGATGCA AAGACAAAGG CACAGCTGGA TGAATGGAAG GCAAAGAAGA AAGCTTCTAA 720  
 45 TGGGAATGCA AGGCCAGAAA CTGTCACTAA TGACGATGAA GAAGCCTTGG ATGAAGAAAC 780  
 AAAGAGGAGA GATCAGATGA TTAAAGGGGC TATTGAAGTT TTAATTCTGT AATACTCCAG 840  
 50 TGAGCTAAAT GCCCCCTCAC AGGAATCTGA TTCTCACCCC AGGAAGAAGA AGAAGGAAAA 900  
 GAAGGAGGAC ATTTTCGCA GATTTCCAGT GGGCCCACTG ATCCCTTATC CACTCATCAC 960  
 TAAGGAGGAT ATAAATGCTA TAGAAATGGA AGAAGACAAA AGAGACCTGA TATCTCGAGA 1020  
 55 GATCAGCAAA TTCAGAGACA CACATAAGAA ACTGGAAGAA GAGAAAGGCA AAAAGGAAAA 1080  
 AGAAAGACAG GAAATTGAGA AAGAACGGAG AGAAAGAGAG AGGGAGCGTG AAAGGGAACG 1140  
 60 AGAAAGGCGA GAACGGGAAC GAGAAAGGGA AAGAGAACGT GAACGAGAAA AGGAGAAAGA 1200

ACGGGAGCGG GAACGAGAAC GGGATAGGGA CCGTGACCGG ACAAAGAGA GAGACCGAGA 1260  
5 TCGGGATCGA GAGAGAGATC GTGACCGGGA TAGAGAAAGG AGCTCAGATC GTAATAAGGA 1320  
TCGCATTCTGA TCAAGAGAAA AAAGCAGAGA TCGTGAAAGG GAACGAGAGC GGGAAAGAGA 1380  
GAGAGAGAGA GAACGAGAGC GAGAACGAGA ACGGGAGCGA GAGAGAGAAG C 1431

10

(2) INFORMATION FOR SEQ ID NO: 22:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2539 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

GGGTGCAGGA GTGCCACCCC CAGGGCCCTG TCAACCTCTC TTTCTCCTC CATGGCTGTC 60  
25 TGCTGCGTA TCTGTCTCTG AGAATCCTCG GGGCGGTCAG GGGATGTCAG GAGGGGAAGG 120  
AGCCGCCCTC CCTATCTTGC TGCTCTCTTT GGCACCTCAG GGCACCTTCC ATGGAGCCAG 180  
ACCGGGTGGA GGGGCTTCTG GGATTGGTG TCTGCTGCTG CCAGAGCAGG AACCCCCAGT 240  
30 CTAGGACTTG GGCATTTTAA CAGGGAGAAA GTAGTGGCTT CCCTTTTCTC TCTCTCTCC 300  
TTTTTCCCTT TAAGCCACCA GATTCAGGTC ATGCCAAAAG CTCTCTGGTT GTAACCTGGA 360  
35 GACATGTGGA GGGGAATGGC GATGGGATTA TAGGACTCTC CCCATCTCGG GCCCTGACCC 420  
TGACCCCTGC CACCAACCCA AAGACAGCTG GTGGGTTTCC CCTTGGAGAM AATCCTGCGT 480  
TTGCCTGGGC CGGCCCTGGC TGCCCTCAGC TTTCGCTGAT CTGCCCGGCC TGGAGCCTCC 540  
40 CATCACCCCG CTTCTTGTG GGCCTCAGGC ACTGGTTACC AGAAGGGGGT CTGGGTCTGC 600  
TCAGGAATCA TGTTTTGTAG CACCTCCTGT TGGAGGGGTG GAGGGATGTT CCCCTGAGCC 660  
45 AGGCTGAGAC TAGAACCCCA TCTTCCTGA GCCAGGCTGA GACTAGAACC CCATCTTCCC 720  
CACCACGCCA CCCCTGTGST KGCTACAGGA GCACAGTAGT GAAGGCCTGA GCTCCAGGTT 780  
TGAAAGACCC AACTGGAGCG TGGGGCGGGC AGGCAGGGT TAGTGAAAGG ACACTTCCAG 840  
50 GGTTAGGACA GAGCATTTAG CCTTCTGGAA GAACCCCTGC CTGGGGTGGG ACTGTGCAGG 900  
CCAGAGAAGG TGGCATGGGC CTGAACCCAC CTGGACTGAC TTCTGCACTG AAGCCACAGA 960  
55 TGGAGGGTAG GCTGGTGGGT GGGGGTGGTT CGTTCTCTAG CCGGGGCAGA CACCCAGCTG 1020  
GCTGGGTCTT TCCTCAGCT TGCTCCTCC TGTCCCAAC CCTTCTCTT CCTCCTGCTT 1080  
60 GCGGACTGCT GGTCCCTCT CCTTCCTCC TTCCAGCTGT TTCTAGTTAC CACCTACCCC 1140

TGGGCCGTGG ACTGATCAGA CCAGCATTC AATAAAAGT TTGTTCCAAG TTGACAGTGT 1200  
GGTGCTCCCT GCCCAGCCCC TCCAGGTGGA GGTGCTGCCA CGGGAACGCA GTTGCTCTGC 1260  
5 CTGCCCTGGG CCCCTGGCGA CANTGGGAGC AGGGCAGTGC TGTGAGGAGC CCAGCTTTCC 1320  
CAGTCAGGCA GGCATGGCTT CCGTGTTCAG GCTCCCTCAC CAGCTGGTGA CACGGGACAA 1380  
10 GCTTACAAAC CTCTCTGAA CCTCAGTTTT CTCATTTACA AGAGGCAAAG CATCCATCAC 1440  
CTTGTTGGA TTCARAGAAT GTRAGGCCCT GGGGTGTCCT ACACAAGGGA AAGGCTTGCT 1500  
CAGTGAGCGG TCTGCACACC GTTAGCCACC CTGCCACCTC TGTGCCCTGG GCAGGCTCCA 1560  
15 AAGGAAAGCT CTGGCTGGGA CTGCCRGAG TCTCACACGC TCCTGTTGAC ATTCCCAGCA 1620  
GCVGCCCTG AGGTCGATGT TTGTTCTGTT TTTCTTTTC TTTTGTGAGA CGGAGTCTCG 1680  
CTGTGTTGCC AGGCTGGAGT GCAGTGGTGT GATCTCTGCT CACTGCAACC TCCGCTGCC 1740  
20 AGTTTCAAGT GATTCTCTGC CTCAGCCTTC TGAGTAGCTG GGACTACAGG TGCACGCCAC 1800  
CACGCCAGC TAACTTTTTG TATTTWAGTA GAGACAGGT TCGCCATGT CGGCCAGGT 1860  
25 GGTCTTGATC TCCTGACCTC ATGATCCACC CGCTCAGCC TCCCAAAGTG CTGGGATTAC 1920  
AGGTATGAGC CACCGCACCG GGCCTGTCTT ATTTTCTAG TTAAGGAAC TGAAGCTCAG 1980  
ARAGGTGTCA CCAGCARGTG TTCATTCCCA TGCCAGCCTT GCGCCCGGC TTTTCCAGG 2040  
30 CAGGCTCTG CGTGCCCACT GGCTCCAGCC TGGTCTCTG TCTCTTGGCT GCTTCACTCC 2100  
TGCTCTTTGT CCCGACTCTG GGCCTGCTTA CAGGGGCCAC TACCTGCTGG TGCCTCCATA 2160  
35 ACAAGCGTCT GCGGTGAGA CCCCTGGCAT GGCAGGGGCT TTGGGGTCTG GTTCCACAA 2220  
GGCTAGCCA TGGCAGAACC TCGTTTATT TTAATCTTT GCGCCTACAA ACAACAGCA 2280  
GTACTTGCCA GAACCATTTCT TGGGATTGAG GAGCTGGGC GACTGCCTTG GCCTCTGGCC 2340  
40 GCACCCAGGA GGGTGGGGTT GGATCTGTGT AGTTGCCAGG CCCACACCTG CCAGCAGGGG 2400  
GCTGACTGGA TCCATGCTTT ACTGTGTTTA ATGGGGTAA CAGGGGTCC TACAGCCCTC 2460  
45 CCAGYTAAAM ATTTGGAACA AAACACCAGC CCTTTGTAG TGGATGCAGA ATAAAATTGT 2520  
TAATCCAATC AAAAAAAAAA 2539

50

## (2) INFORMATION FOR SEQ ID NO: 23:

- (i) SEQUENCE CHARACTERISTICS:  
55 (A) LENGTH: 1041 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

- 60 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

|    |   |      |
|----|---|------|
|    | TCGACCCACG CGTCCGCCCA CGCGTCCGCC CACGGGTCCG GGCGCAGGAC GTGCACTATG | 60   |
| 5  | GCTCGGGGCT CGCTGCGCGG GTTGCTGCGG CTCTCTGTGC TGGGGCTCTG GCTGGCGTTG | 120  |
|    | CTGCGCTCCG TGGCGGGGA GCAAGCGCCA GGCACCGCCC CCTGCTCCCG CGGCAGCTCC  | 180  |
|    | TGGAGCGCGG ACCTGGACAA GTGCATGGAC TGGCGGTCTT GCAGGGGCG ACCGCACAGC  | 240  |
| 10 | GACTTCTGCC TGGGTGCGC TGCAGCACCT CCTGCCCCCT TCCGGCTGCT TTGGCCATC   | 300  |
|    | CTTGGGGGCG CTCTGAGCCT GACCTTCGTG CTGGGGCTGC TTTCTGGCTT TTTGGTCTGG | 360  |
| 15 | AGACGATGCC GCAGAGAGAG AAGTTCACCA CCCCATAGA GGAGACCGGC GGAGAGGGCT  | 420  |
|    | GCCAGCTGT GGCCTGATC CAGTGACAAT GTGCCCCCTG CCAGCCGGGG CTCGCCCACT   | 480  |
|    | CATCATTCAT TCATCCATTG TAGAGCCAGT CTCTGCCTCC CAGACGCGGC GGGAGCAAGC | 540  |
| 20 | TCCTCCAACC ACAAGGGGGG TGGGGGGCGG TGAATCACCT CYGAGGCCTG GGGCCAGGGT | 600  |
|    | TCAGGGGAAC TTCCAAGTG TCTGGTTGCC CTGCCTCTGG CTCCAGAACA GAAAGGGAGC  | 660  |
| 25 | CTCAGCTGG CTCACACAAA ACAGCTGACA CTGACTAAGG AACTGCAGCA TTTCACAGG   | 720  |
|    | GGAGGGGGT GCCCTCCTTC CTAGAGGCC TGGGGGCCAG GCTGACTTGG GGGGCAGACT   | 780  |
|    | TGACACTAGG CCCCACTCAC TCAGATGTCC TGAAATTCCA CCACGGGGT CACCCTGGGG  | 840  |
| 30 | GGTTAGGGAC CTATTTTAA CACTAGGGG CTGGCCCACT AGGAGGGCTG GCCCTAAGAT   | 900  |
|    | ACAGACCCCC CCAACTCCCC AAAGCGGGA GGAGATATTT ATTTTGGGA GAGTTTGGAG   | 960  |
| 35 | GGGAGGGAGA ATTTATTAAT AAAAGAATCT TTAACTTAA AAAAAAAAAA AAAAAAGGC   | 1020 |
|    | GGCCGCTCTA GAGGATCCCT C   | 1041 |

40

(2) INFORMATION FOR SEQ ID NO: 24:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 1962 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

|    |   |     |
|----|---|-----|
| 50 | ACCCACGCGT CCGGTACAAA ACACAGTTTT ATTCTATGAA AATTTTGAGA TTATTAGAAA | 60  |
|    | CATTAGATTT AGGGTTGCAT ATTAAAACT ATATCCATTT TGCCTTATTA TTTAGTGTCT  | 120 |
| 55 | CACTCAGGAT ATAACACACT ATAATAGAAA ATGTAGACTT CAGAATCAGG TATATTGAG  | 180 |
|    | ATGGTTTGTA TACTGGTTCT GACACTTGTT AGCTATTCAT CTTTGGTAAA TTCCCATTA  | 240 |
| 60 | CCCTTTGTC ACCTATWGT GGGGATCAGT GCATAGTGTG TGTWAAGCAT TTAATACCTG   | 300 |

|    |   |      |
|----|---|------|
|    | GCAAGTGTTC AGCAAATTTT TTGTTCTATA TATTTATTAT TTGATTATG GCCCTGAGGA  | 360  |
|    | GTAGGTGTTT GTTTGTTTGT TTGTTTGT TT AGTTTATTT CTCATCTCCT CAGGAACACA | 420  |
| 5  | AATGAACTT GGATATTGTT ATGGTGCTTT TNATAATATA TTTATTATTT TCAGCAATTN  | 480  |
|    | ATTCTTGTTA AAACAATTTT TTATGACAAG TTAATCATCT TCAATGGTGA GAAGAAATCT | 540  |
| 10 | AGCTCAGAAT AATATATTTT TAGTGTGTTT ATCTCTGGAT ACTCATTTTG CTCATTGCCA | 600  |
|    | CGTAAAGTAA AAAAATACAT AAATTAGCTT ATTCCAATGT AATATCTTCA GGATAGTCAT | 660  |
|    | GGGCAAGGAA TTAATCACAT TAAGAGATAA CTGCAACTAA GCACTATTTG AGGTGACTTC | 720  |
| 15 | TGTGGAAAAA AAATTAATYC TTTACCATTG CAGCGTTCTG CCCTAGGTCC AAATGTTACC | 780  |
|    | AAAATCACTC TAGAATCTTT TCTTGCTGG AAGAAAAGGA AAAGACAAGA AAAGATTGAT  | 840  |
| 20 | AAACTTGAAC AAGATATGGA AAGAAGGAAA GCTGACTTCA AAGCAGGGAA AGCACTAGTG | 900  |
|    | ATCAGTGGTC GTGAAGTGT TGAATTCGT CCTGAACTGG TCAATGATGA TGATGAGGAA   | 960  |
|    | GCAGATGATA CCCGCTACAC CCAGGGAACA GGTGGTGATG AGGTTGATGA TTCAGTGAGT | 1020 |
| 25 | GTAAATGACA TAGATTTAAG CCTGTACATC CCAAGAGATG TAGATGAAAC AGGTATTACT | 1080 |
|    | GTAGCCAGTC TTGAAAGATT CAGCACATAT ACTTCAGATA AAGATGAAAA CAAATTAAGT | 1140 |
| 30 | GAAGCTTCTG GAGGTAGGGC TGAAAATGGT GAAAGAAGTG ACTTGAAGA GGACAACGAG  | 1200 |
|    | AGGGAGGGAA CGGAAAATGG AGCCATTGAT GCTGTTCTG TTGATGAAAA TCTTTTCACT  | 1260 |
|    | GGAGAGGATT TGGATGAACT AGAAGAAGAA TTAAATACAC TTGATTTAGA AGAATGACAC | 1320 |
| 35 | CAAACACATC GCTGAAAAAA TTAAGTCAGC TCAGCACGAG TTGAAATTGA CTACATTAAT | 1380 |
|    | TTCTTTCCAC CTAGAATCAA CAGGATGTTT ATTTCCTATG CTGATTCTGG AGGAGTTAAC | 1440 |
| 40 | CTCCTGCAAA AAAGGCATCT TGTCCCTACA TCTTCTCTTC TGACTTTGGC TACATCTCAT | 1500 |
|    | AGTAAGTTCA GAGTAGTTCA TGATAAATTG AAAATATAAT GGTCATTGCA GAAAATGATT | 1560 |
|    | GATGTTGTAA CTGTCCACCC AAGTAAGAAG TGTATCTGCC TTTCCATCTT TTGGTTTCA  | 1620 |
| 45 | TTTGGGCATG TGCTATTACC AGAAACAACA AACTTATAIT TAAAATACCC TTCATTGAC  | 1680 |
|    | ACAGTTTTTA ATGAGTGATT TAATTTCTC TGTATTTGTA TGTTTAGAAG ACTGCCTAAA  | 1740 |
| 50 | ACATGAGCAC TGTACTTCAT AAAGGAAACG CGTATGCAGA TTCAGTATTG TGTATCTTTG | 1800 |
|    | GACAATTAGA TGGACATTTA AAATGGAAC TCTTTTATCT GACAGGATCA GCTACAATGC  | 1860 |
|    | CCTGTGTAA ATGTTTTAAA AGTTTCCCTT TTCTTTTTTG CCAATAAAGT TGTAATAAAA  | 1920 |
| 55 | GACCATCATA CATTAATAAT CAAAAA AAAA AAAA AA                         | 1962 |

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1228 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

10 GGCTGCCCAG GCCCCGCACT GGAAGAGCCT CCAGCAGCAA GATGTGACCG YTGTCGCGAT 60  
 GAGCCCCAGC AGCCACTCCC CAGAGGGGAG GCCTCCACCT CTGCTGCCTG GGGGTCCAGT 120  
 GTGTAAGGCA GCTGCATCTG CACCGAGCTC CCTCCTGGAC CAGCCGTGCC TCTGCCCCGC 180  
 15 ACCCTCTGTC CGCACCCTG TTGCCCTGAC AACGCCGGAT ATCACATTGG TTCTGCCCCC 240  
 TGACATCATC CAACAGGAAG CGTCACCCTG AGGGAGGAGA CAGAAGCCTG GGCCAGGTGA 300  
 20 ACAGTGGTAT AGCAGCCACT CCAGCCTCTG CTGCAGCAGC CACCCTGGAT GTGGCTGTTT 360  
 GGAGAGGCCT GTCCACGGA GCCCAGAGGC TGCTGTGCGT GGGCCTGGGA CAGCTGGACC 420  
 25 GGCCTCCAGA CCTCGCCCAT GACGGGAGGA GTCTGTGGCT GAACATCAGG GGCAAGGAGG 480  
 CGGCTGCCCT ATCCATGTTT CATGTCTCCA CGCCACTGCC AGTGATGACC GGTGGTTTCC 540  
 TGAGCTGCAT CTGGGGCTTG GTGCTGCCCC TGGCCTATGN TTCCAGCCTG ACCTGGTGCT 600  
 30 GGTGGCGCTG GGGCCTGCCA NTGCCTGCAG GGCCCCCAG CTGCACTCCT GGCTGCAATG 660  
 CTTGCGGGGC TGGCAGGGG CCGAGTCCTG GCCCTCCTGG AGGAGAACTC CACACCCAG 720  
 35 CTAGCAGGGA TCCTGGCCCG GGTGCTGAAT GGAGAGGCAC CTCCTAGCCT AGGCCCTTCC 780  
 TCTGTGGCCT CCCCAGAGGA CGTCCAGGCC CTGATGTACC TGAGAGGGCA GCTGGAGCCT 840  
 CAGTGAAGA TGTGTCAGTG CCATCCTCAC CTGGTGGCTT GAAATCGGCC AAGGTGGGAG 900  
 40 CATTTCACAC GCAGAAATGA CACCGCACGC CAGCGCCCCG CGGCCGCGAT CCGGACCCCA 960  
 AGCCCCACGC TCCTCGACT CTGGGGCAGC GAACCCCGCC CACTCCCAAT CCGCGCGCCC 1020  
 45 CGCCCTCTCC CACCCGTGCT TCCCCCGCTC CACCCCTCAC CTCACCTCGC CCCSGCCCCA 1080  
 CCCATCGCGC CCGGGCCCGT CCCATCGAGG CCCATGCAAC CCACGCTCGG TYCCGTTCCG 1140  
 GCCCCTGCGC TCKGCTKNS TTCGCTCCCC GCCCTGCGC CGTTAGTAAA CATCGCTCAA 1200  
 50 ACGAAAAAAA AAAAAAAAAA AAACCTGA 1228

## 55 (2) INFORMATION FOR SEQ ID NO: 26:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1340 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

|    |   |      |
|----|---|------|
| 5  | AATTCGGCAG AGAGATGGCC GCGCCCGTGG ATCTAGAGCT GAAGAAGGCC TTCACAGAGC | 60   |
|    | TTCAAGCCAA AGTTATTGAC ACTCAACAGA AGGTGAAGCT CGCAGACATA CAGATTGAAC | 120  |
|    | AGCTAAACAG AACGAAAAAG CATGCACATC TTACAGATAC AGAGATCATG ACTTTGGTAG | 180  |
| 10 | ATGAGACTAA CATGTATGAA GGTGTAGGAA GAATGTTTAT TCTTCAGTCC AAGGAAGCAA | 240  |
|    | TTACAGTCA GCTGTTAGAG AAGCAGAAAA TAGCAGAAGA AAAAATTAAA GAACTAGAAC  | 300  |
| 15 | AGAAAAAGTC CTACCTGGAG CGACGTTAAA GGAAGCTGAG GACAACATCC GGGAGATGCT | 360  |
|    | GATGGCACGA AGGGCCCGT AGGGAGCCTC TCTGGGAAGC TCTTCCTCCT GCCCTCCCA   | 420  |
|    | TTCTTGGTGG GGGCAGAGGA GTGTCTGCAG GGAAACAGCT TCTCCTCTGC CCCGATGGAT | 480  |
| 20 | GCTTTATTTG GATGGCCTGG CAACATCACA TTTTCTGCAT CACCTGAGC CCCATTTGCT  | 540  |
|    | TCCCAGCCCT GGAGTTTTTA CCCGGCTTTG CTGCCACCTC TGCCAGGAC ACKCTTCCCT  | 600  |
| 25 | CTCGGATGT GTGATGAAC CCCAGGAGAG GGAAGATGGG AGCCAGGGCA AGATAGGAAG   | 660  |
|    | CTCTGCCTGA GCTTTCCACT AGGCACGCCA GCCAGACCAA TAAAAAGCGT CTGTCCACT  | 720  |
|    | CTGCTAAGCC TGGTTTCTT GAGCAGAGGG ATGGAACAGA GGGTGAGAGA GGCAGTGGCC  | 780  |
| 30 | GTCTCCACCT CAGCTCCTGC TCCCTCTGCA TCAGAGCCCT TCCTTTCTTG GGGGATGGGC | 840  |
|    | CTTGCCNTCT TCTCTTTTCC CTTCTGTAC CTTTGACTAA CGCTCAGCTT CCGGGCCTGC  | 900  |
| 35 | ATGCAGTAGA CAGAAGAGGA AGAAAGAACA GATGTTTACA GCTGAATCTC AGTGAACAGA | 960  |
|    | ATAGCAGTCC CTGGATGGCA GTCTGCCTAA AGATTCTTT CCCTGCCTTC TCCCATACAT  | 1020 |
|    | TCCAAAAGGA AGTTCAACAG TAAGCAGCAC CTCCAAGACT GTCTCCTTTT GCCARTATC  | 1080 |
| 40 | ATAAGATGGA CGCCATAATC CTGAGGCCTC CTAGAGGCTG AGGGGGCAAC GGTGTGATCC | 1140 |
|    | AGCTGGCTCA TCCAGCCAG GTGGGCCAAT TATTCAATTT TCAAGAATTT TGTGCAAGC   | 1200 |
| 45 | CAGTTGTCAA ACACAGCCAT TATAATTATG TAAATTTGCA AATTATGTTA AAAACAAGGA | 1260 |
|    | CAATAAATAT TCAAATGCA TCCCTAAWWA AAAAAAAAAA AANGGGNGGC CGCNCATAGGG | 1320 |
| 50 | GATCCAAGCT TACGTACGG  | 1340 |

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 806 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

5 ACCTTCTTCC ATGTTTAGTC CCTTGGGCTC TGCTACCCCTC CTGCTGGAGG TGAGAGCATC 60  
 CTGTGTGCAA CCAGAGATGC CCTCTGGCTT TCAGACCTGC CTGCTTTTCA CCCTCAGCCC 120  
 TTTCTCACTC AGCAAAATTG TGGGGGTCCC TAGTCAGCAG CTCCTTGGGC AGCTCTCTGA 180  
 10 GCAAGGTGGT CTCTGTGGTC ATGAAGGAGA GCCGGCTAGG ACAGTGCCGG AAATCAGCT 240  
 GCCTCTCCCC TTCAACTCAG CTGGCCCCCC GCACCTGAAG TGCACAGGAG CCGGAAGAG 300  
 15 AGTCTGGAGC CCACCCCGGA GGGCAGCACA GGAGGTGTCT CTGCAGCTGG TGTCTTGCCA 360  
 CCCCCTGCAGG CAGCACACGT CCCGGGCATT CTCCTTAGCC ACAGACAGAA CAGCCAGTGC 420  
 CAGAGTCTGC TGTCGTTCCT CTTAAGCAC ACTCATTCAC CACACCCGAG GAGGCCAGAG 480  
 20 GTGCAGGGAG CATGGGCTGT CGCTTCCCTT TTAAGCACAC TCATTACCA CACCCGAGGA 540  
 GGCCAGAAGT GCAGGGAGCA TGGGCTGGGT GCACCTCCGC AGGAGAGAAG GCTGAGCCAC 600  
 25 CGCCGTCCCG GGAGCCCGGC TCCCAGGCCT CTCGTTTTCC CCTACCTCCC TAAGACTTTT 660  
 CTGTCACTCT CTGGCCATTG AAAGGCTTCT GTTCCTTAAA GTGCTGTIAC ACTCTCCTTT 720  
 CCCAGGATGC AGCAAGCCAA AACAGTACCA CTGCACGTCA GCCTGGGTGA CAGAGTGAGA 780  
 30 CCCTATCTTA AAAAAAAAAA AAAAAA 806

(2) INFORMATION FOR SEQ ID NO: 28:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 696 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

45 GAGTTCCTNA CCGGTGGCG NCCGTTTITAG AAATTAGTGG ATCCCCCGG GCTGGCAGGG 60  
 AATTGGGCAC GAGCACAGAG GAAAGCGGGT GCCCGGCATG GCCATCCTGA TGTGTCTGGC 120  
 50 GGGATCCCCA TGCACCTTGT CTTTCTCCAC TGATACTGGC AGCTCGGCTC CTGGACCCAA 180  
 GATCCCTTGA GTGGAATTCT GCAGTGCAAG AGCCCTTCGT GGGAGCTGTC CCATGTTTCC 240  
 ATGGTCCCCA GTCTCCCTC CACTTGGTGG GGTACCAAC TACTCACCAG AAGGGGGCTT 300  
 55 ACCAAGAAAG CCTTAAAAAG CTGTTGACTT ATCTGCGCTT GTTCCAATC TTATGCCCCC 360  
 AACCTGCCCT ACCACCACCA CGCGCTCAGC CTGATGTGTT TACATGGTAC TGTATGTATG 420  
 60 GGAGAGCAGA CTGCACCTC CAGCAACAAC AGATGAAAGC CAGTGAGCCT ACTAACCGTG 480

CCATCTTGCA AACTACACTT TAAAAAAAC TCATTGCTTT GTATTGTAGT AACCAATATG 540  
TGCAGTATAC GTTGAATGTA TATGAACATA CTTTCCTATT TCTGTTCTTT GAAAATGTCA 600  
5 GAAATATTTT TTTCTTTCTC ATTTTATGTT GAACTAAAAA GGATTAAAAA AAAAATCTCC 660  
AGAMAAAAAA AAAAAA AAAATTACTGC GGTCCG 696

10

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:  
15 (A) LENGTH: 1007 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AATTGGGCAC GAGGAAAAA TACCATTGTG GTATGATACC CAATTGGAT CTCAATTGG 60  
ATAGAGATTT GGTGCTTCCA GATGTRAGTT ATCAGGTGGA ATCCAGTGAG GAGGATCAGT 120  
25 CTCAGACTAT GGATCCTCAA GGACAACTC TGCTGCTTTT TCTCTTTGTG GATTTCACA 180  
GTGCAATTCC AGTCCAGCAA ATGGAAATCT GGGGAGTCTA TACTTTGCTC ACAACTCATC 240  
30 TCAATGCCAT CCTGTGGAG AGCCACAGTG TAGTGCAAGG TTCCATCCAA TTCACTGTGG 300  
ACAAGGTCTT GGAGCAACAT CACCAGGCTG CCAAGGCTCA GCAGAACTA CAGGCCTCAC 360  
TCTCAGTGGC TGTGAACTCC ATCATGAGTA TTCTGACTGG AAGCACTAGG AGCAGCTTCC 420  
35 GAAAGATGTG TCTCCAGACC CTTCAAGCAG CTGACACACA AGAGTTCAGG ACCAACTGC 480  
ACAAAGTATT TCGTGAGATC ACCCAACACC AATTTCTTCA CCACTGCTCA TGTGAGGTGA 540  
40 AGCAGCTAAC CCTAGAAAAA AAGGACTCAG CCCAGGCAC TGAGGACGCA CCTGATAACA 600  
GCAGCCTGGA GCTCCTAGCA GATACCAGCG GGCAAGCAGA AAACAAGAGG CTCAAGAGGG 660  
GCAGCCCCCG CATAGAGGAG ATGCGAGCTC TCGCTCTGCG CAGGGCCCCG AGCCCGTCAG 720  
45 AGGCCGCCCC GCGCCGCCCC GAAGCCACCG CGGCCCCCT CACTCCTAGA GGAAGGGAGC 780  
ACCGCGAGGC TCACGGCAGG GCCCTGGCGC CGGGCAGGGC GAGCCTCGGA AGCCGCCTGG 840  
50 AGGACGTGCT GTGGCTGCAG GAGGTCTCCA ACCTGTCAGA GTGGCTGAGT CCCAGCCCTG 900  
GGCCCTGAGC CGGGTCCCTT TNCGCAAGCG CCCACCGATC CGGARGCTGC GGGCAGCCGT 960  
55 TATCCCGTGG TTTAATAAAG TGCGCGCGCG TCACCAAAAA AAAAAA 1007

60

(2) INFORMATION FOR SEQ ID NO: 30:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2017 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

5 AATTCCGGCAC GAGCGGATCC GTTGC GGCTG CAGCTCTGCA GTCGGGCGGT TCCTTCGCCG 60  
 10 CCGCCAGGGG TAGCGGTGTA GCTGCGCAGC TCGCGGCGC TACCGCACCC AGGTTCGGCC 120  
 CGTAGCGTCT GGCAGCCCGG CGCCATCTTC ATCGAGCGCC ATGGCCGCAG CCTGCGGGCC 180  
 15 GGGAGCGGCG GGTACTGCTT GCTCCTCGGC TTGCATTTGT TTCTGCTGAC CGCGGGCCCT 240  
 GCCTGGGCTG GAACGACCCT GACAGAATGT TGCTGCGGGA TGTAAGAGCT CTTACCCTCC 300  
 20 ACTATGACCG CTATACCACC TCCCGCAGCT GGATCCCATC CCACAGTTGA AATGTGTTGG 360  
 AGGCACAGCT GGTGTGTGATT CTTATACCCC AAAAGTCATA CAGTGTGAGA ACAAAGGCTG 420  
 GGATGGGTAT GATGTACAGT GGAATGTAA GACGGACTTA GATATTGCAT ACAAATTTGG 480  
 25 AAAAAGTGTG GTGAGCTGTG AAGGCTATGA GTCCTCTGAA GACCAGTATG TACTAAGAGG 540  
 TTCTGTGGC TTGAGTATA ATTTAGATTA TACAGAACTT GGCCTGCAGA AACTGAAGGA 600  
 30 GTCTGAAAG CAGCAGGCT TTGCCTCTTT CTCTGATTAT TATTATAAGT GGTCTCGGC 660  
 GGATTCCTGT AACATGAGTG GATTGATTAC CATCGTGGTA CTCCTTGGGA TCGCCTTTGT 720  
 AGTCTATAAG CTGTTCTGA GTGACGGCA GTATTCTCTT CCACCGTACT CTGAGTATCC 780  
 35 TCCATTTTCC CACCGTTACC AGAGATTCAC CAACTCAGCA GGACCTCTC CCCCAGGCTT 840  
 TAAGTCTGAG TTCACAGGAC CACAGAATAC TGGCCATGGT GCAACTTCTG GTTTTGGCAG 900  
 TGCTTTTACA GGACAACAAG GATATGAAAA TTCAGGACCA GGGTTCTGGA CAGGCTTGGG 960  
 40 AACTGGTGA ATACTAGGAT ATTTGTTTGG CAGCAATAGA GCGGCAACAC CCTTCTCAGA 1020  
 CTCGTGGTAC TACCCGTCTT ATCCTCCCTC CTACCTGGC ACGTGAATA GGGCTTACTC 1080  
 45 ACCCCTTCAT GGAGGCTCGG GCAGCTATTC GGTATGTTCA AACTCAGACA CGAAAACCAG 1140  
 AACTGCATCA GGATATGGTG GTACCAGGAG ACGATAAAGT AGAAAGTTGG AGTCAAACAC 1200  
 50 TGGATGCAGA AATTTTGGAT TTTTCATCAC TTTCTCTTTA GAAAAAAGT ACTACCTGTT 1260  
 AACAAATGGG AAAAGGGGAT ATTCAAAAGT TCTGTGGTGT TATGTCCAGT GTAGCTTTTT 1320  
 GTATCTTATT ATTTGAGGCT AAAAGTTGAT GTGTGACAAA ATACTTATGT GTGTATGTC 1380  
 55 AGTGTAACAT GCAGATGTAT ATTGCAGTTT TTGAAAGTGA TCATTACTGT GGAATGCTAA 1440  
 AAATACATTA ATTTCTAAAA CCGTGATGTC CCTAAGAAGC ATTAAGAATG AAGGTGTTGT 1500  
 60 ACTAATAGAA ACTAAGTACA GAAAATTTCA GTTTTAGGTG GTTGTAGCTG ATGAGTTATT 1560

ACCTCATAGA GACTATAATA TTCTATTTGG TATTATATTA TTTGATGTTT GCTGTTCTTC 1620  
AAACATTTAA ATCAAGCTTT GGAATAATTA TGCTAATTTG TGAGTTCTGA TCACTTTGA 1680  
5 GCTCTGAAGC TTTGAATCAT TCAGTGGTGG AGATGGCCTT CTGGTAACTG AATATTACCT 1740  
TCTGTAGGAA AAGGTGGAAA ATAAGCATCT AGAAGGTTGT TGTGAATGAC TCTGTGCTGG 1800  
CAAAAATGCT TGAACCTCT ATATTCTTT CGTTCATAAG AGGTAAAGGT CAAATTTTTC 1860  
10 AACAAAAGTC TTTTAATAAC AAAAGCATGC AGTTCTCTGT GAAATCTCAA ATATTGTTGT 1920  
AATAGTCTGT TTCAATCTTA AAAAGAATCA ATAAAAACAA ACAAGGGAAA AAAAAAAAAA 1980  
15 AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA 2017

20 (2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

25 (A) LENGTH: 699 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

30 GNGTTTTTTC CAGCCAGGAA GTGACCGNTA CTGCAGCAGC AGANAGATTG GTTGGGTTGG 60  
TTGRAAATGA CYCTGAACAT TTATTTCCAT TGCAATTTCT GTGGCTGAGG AGACTTAAAC 120  
TTTACAAGTA TTATCCTTTT AAGATCATTT TAATTTTAGT TGAGTGCAGA GGGCTTTTAT 180  
35 AACAAACGTG CAGAAATTTT GGAGGGCTGT GATTTTCCA GTATTAAACA TGCATGCATT 240  
AATCTGTCAG TTTATTTTCT CATGTGTAT GTATATATCG CTTTCTCTG CAGCAGGATT 300  
40 TCTCTTTTGA TAAWKCCCTT TAGGGCACAA CTAGTTATCA GTAAGTGAAT GTATCTTAAT 360  
CATTATGGCT GCTTCTGTTT TTTCATTAAC AAAGGTTATT CATATGTTAG CATATAGTTT 420  
CTTTGCACCC ACTATTTATG TCTGAATCAT TTGTCACAAG AGAGTGTGTG CTGATGAGAT 480  
45 TGTAAGTTTG TGTGTTTAAA CTTTTTTTGG AGCGAGGGAA GAAAAAGCTG TATGCATTTT 540  
ATTGCTGTCT ACAGGTTTCT TTCAGATTAT GTTCATGGGT TTGTGTGTAT ACAATATGAA 600  
50 GAATGATCTG AAGTAATTGT GCTGTATTTA TGTTTATTCA CCAGTCTTTG ATTAAATAAA 660  
AAGGAAAACC AGAAAAAAAA AAAAAAAAAA AAAAAAAA 699

55

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

60 (A) LENGTH: 1264 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GGCACGAGGG CACTGTTTCC TCAGTCCATG GCTGAGTACA TCACCGGTGT TTTCTCTCTT 60  
ATTCCTCCCA TCAAGCCTAA AAGGAATCTC TATTGGAGAT ACTGCCATTA GTGTTCCMTT 120  
10 TATAGGTGAG GAACTGAGGC ATAKAGGGTT CCCCAGTTGA ACCAACTGAT AAATAGTAGA 180  
ACTTGGATT TTAATTCAGTC TTGATGCCAG GGATAAGGCT CTTACTTTCT ACCTTAGGCT 240  
15 ATTTCTAGGA AACGCAGGAG AGTGTGAAG GGCAGAGAA AGGGATCCAG TTCCTTTCTG 300  
TCCCGCATCC TAGTCCCTGA GAAGCAAAGA ARAATGTGTG GCTTCTTTTG CTTTGCTTTT 360  
GTTGTCATCC CACACATCTC CAGGGGAMCT GGGCTCTTGA TCTTGGSC TCCTCCCTTTA 420  
20 ACTGTTAAGT GGGAGCARGT AAGGGGTAC AGTAGGGCTG GCCTGGAGTT AGAGGCTTGG 480  
ATGCCITAGC TCCTCTGTCT GCACTCCAGA ACTGCCTGAC TTCATTTCGT ATGTTGTCTT 540  
25 TTGTTTGTAC AATTGATCCA TGTCCAGTC CGTCTCTTCT TCCTTCTTGA TACTTACACT 600  
GCTTCTTTCT GTTGGTTTCC AGTGTTTAAC ACTGTATACA ACAGTGACGA CAACGTGTTT 660  
GTGGGGGCCC CCACGGGCAG CGGGAAGACT ATTTGTGCAG AGTTTGCCAT CCTGCGAATG 720  
30 CTGCTGCAGA GCTCGGAGGG GCGCTGTGYS TWCWTCACCM CCATGGAGGC CCTGGCCAGA 780  
RCAGGTATGA CGTGGCGCTG TGTGATGTA ATTTCCCAAG AAGCATTTCA TCTGTGATTC 840  
35 CGTATGAAGG CTTTCTAAGC CCTGAAATTT GCAGGGTCAT TTCTCAGTT TGTGTATTAA 900  
AGAAAAGCTG CCCCAGCCAA GCGTGGTGGC TCAAGCCTGT AATCCCAGCA CTTTGGGAGG 960  
CCGAGGCGGG CAGATCTCCG GAGATCAGGA GTTCGAGACC AGCCTGGCCA ACATGGTGRA 1020  
40 ACCCTGTCTC TACTAAAAT ACAGAAATTA GCTGGGNGTG GTGGTGTGCG CCTGTAATCC 1080  
CAGCTACTTG GAAGGCTGAG GCAGGAGAAT CGCTTGAACC CGGGAGGCGG AGGTTGCAGT 1140  
45 GAGCCAAGTT CGCACCCTG CACTCCAGCC TGGGCAACAA GAGCGAGACT TCATCTCAAA 1200  
AAAAAAAAA AAAAAGCTGA GGGGGGGCCC GGTACCCAAT TCGCCCTATA GTGATCGTAT 1260  
50 TACA 1264

55 (2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 997 base pairs  
(B) TYPE: nucleic acid  
60 (C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

5 ATGTGAAGTT GTTTTGCAAC CTGGGCTTTT ATACAGAAGA ATACGAATCA CAGGTGTGTG 60  
AGCATCTACT TAATTAATTT GCTTACAGCC GATTTCTGCG TTACTCTGGC ATTACCAGTG 120  
AAAATTGTTG TTGACTTGGG TGTGGCACCT TGAAGCTGA AGATATTCCA CTGCCAAGTA 180  
10 ACAGCCTGCC TCATCTATAT CAATATGTAT TTATCAATTA TCTTCTTAGC ATTTGTCAGC 240  
ATTGACCGCT GTCTTCAGCT GACACACAGC TGCAAGATCT ACCGAATACA AGAACCCGGA 300  
TTTGCCAAAA TGATATCAAC CGTTGTGTGG CTAATGGTCC TTCTTATAAT GGTGCCAAAT 360  
15 ATGATGATTC CCATCAAAGA CATCAAGGAA AAGTCAAATG TGGGTTGTAT GGAGTTTAAA 420  
AAGGAATTTG GAAGAAATG GCATTGCTG ACAAATTTCA TATGTGTAGC AATATTTTTA 480  
20 AATTTCTCAG CCATCATTTT AATATCCAAT TGCCTTGTAA TTCGACAGCT CTACAGAAAC 540  
AAAGATAATG AAAATTACCC AAATGTGAAA AAGGCTCTCA TCAACATACT TTTAGTGACC 600  
ACGGGCTACA TCATATGCTT TGTTCCTTAC CACATTGTCC GAATCCCGTA TACCCTCAGC 660  
25 CAGACAGAAG TCATACTGA TTGCTCAACC AGGATTTTAC TCTTCAAAGC CAAAGAGGCT 720  
ACACTGCTCC TGGCTGTGTC GAACCTGTGC TTIGATCCTA TCCTGTACTA TCACCTCTCA 780  
30 AAAGCATTCG GCTCAAAGGT CACTGAGACT TTTGCCTCMC CTAAAGAGAC CAAGGTYAGA 840  
AAGAAAAATT AAGANGTGA AATAATGGCT AAAAGACAGG NTPTTTGTGG TACCAATTCT 900  
GGGCTTTATG GGACCNATAA GTTATTATAG CTGGGAAGGT AAAAAAAAAA AAAGGGNGGG 960  
35 CGCTCTAGAG GTTCCCGAG GGGCCAGCTT AGGGTGC 997

40

## (2) INFORMATION FOR SEQ ID NO: 34:

## (i) SEQUENCE CHARACTERISTICS:

45

- (A) LENGTH: 1914 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

50

GTGTGAGAGG CCTCTCTGGA AGTGTGCCG GGTGTTGCC GCTGGAGCCC GGTGAGAG 60  
GACGAGGTGC CGTGCTTGG AGAATCTTCC GCTGCCGTCG GCTCCCGAG CCCAGCCCTT 120  
55 TCCTAACCCA ACCCAACCTA GCCCAGTCCC AGCCGCCAGC GCCTGTCCCT GTCACGGACC 180  
CCAGCGTTAC CATGCATCCT GCCGTCTTCC TATCCTTACC CGACCTCAGA TGCTCCCTTC 240  
60 TGCTCCTGGT AACTTGGGTT TTTACTCCTG TAACAACTGA AATAACAAGT CTTGATACAG 300

|    |   |      |
|----|---|------|
|    | AGAATATAGA TGAAATTTTA AACAAATGCTG ATGTTGCTTT AGTAAATTTT TATGCTGACT  | 360  |
|    | GGTGTGCTTT CAGTCAGATG TTGCATCCAA TTTTTGAGGA AGCTTCCGAT GTCATTAAGG   | 420  |
| 5  | AAGAATTTCC AAATGAAAAT CAAGTAGTGT TTGCCAGAGT TGATTGTGAT CAGCACTCTG   | 480  |
|    | ACATAGCCCA GAGATACAGG ATAAGCAAAT ACCCAACCCT CAAATTGTTT CGTAATGGGA   | 540  |
| 10 | TGATGATGAA GAGAGAATAC AGGGGTCAGC GATCAGTGAA AGCATTGGCA GATTACATCA   | 600  |
|    | GGCAACAAAA AAGTGACCCC ATTCAAGAAA TTCGGGACTT AGCAGAAATC ACCACTCTTG   | 660  |
|    | ATCGCAGCAA AAGAAATATC ATTGGATATT TTGAGCAAAA GGA CTGCGAC AACTATAGAG  | 720  |
| 15 | TTTTTGAACG AGTAGCGAAT ATTTTGCAATG ATGACTGTGC CTTCCTTTCT GCATTTGGGG  | 780  |
|    | ATGTTTCAAA ACCGGAAGA TATAGTGGCG ACAACATAAT CTACAAACCA CCAGGGCATT    | 840  |
| 20 | CTGCTCCGGA TATGGTGATC TTGGGAGCTA TGACAAATTT TGATGTGACT TACAATTGGA   | 900  |
|    | TTCAAGATAA ATGTGTTTCT CTGTGCCGAG AAATAACATT TGAAAATGGA GAGGAATTGA   | 960  |
|    | CAGAAGAAGG ACTGCCCTTT CTCATACTCT TTCACATGAA AGAAGATACA GAAAGTTTAG   | 1020 |
| 25 | AAATATTCCA GAATGAAGTA GCTCGGCAAT TAATAAGTGA AAAAGGTACA ATAACTTTT    | 1080 |
|    | TACATGCCGA TTGTGACAAA TTTAGACATC CTCTTCTGCA CATACAGAAA ACTCCAGCAG   | 1140 |
| 30 | ATTGTCTGT AATCGCTATT GACAGCTTTA GGCATATGTA TGTGTTTGGA GACTTCAAAG    | 1200 |
|    | ATGTATTAAT TCCTGGAAAA CTCAAGCAAT TCGTATTTGA CTTACATTCT GGAAAACGTC   | 1260 |
|    | ACAGAGAATT CCATCATGGA CCTGACCCAA CTGATACAGC CCCAGGAGAG CAAGCCCAAG   | 1320 |
| 35 | ATGTAGCAAG CAGTCCACCT GAGAGCTCCT TCCAGAACT AGCACCAGT GAATATAGGT     | 1380 |
|    | ATACTCTATT GAGGGATCGA GATGAGCTTT AAAAAGTTGA AAAACAGTTT GTAAGCCTTT   | 1440 |
| 40 | CAACAGCAGC ATCAACCTAC GTGGTGGAAA TAGTAAACCT ATATTTTCAT AATTCTATGT   | 1500 |
|    | GTATTTTTAT TTTGAATAAA CAGAAAGAAA TTTTGGGTTT TTAATTTTTT TCTCCCGAC    | 1560 |
|    | TCAAAATGCA TTGTCATTTA ATATAGTAGC CTCTTAAAAA AAAAAAAAC CTGCTAGGAT    | 1620 |
| 45 | TTAAAAATAA AAATCAGAGG CCTATCTCCA CTTTAAATCT GTCCTGTAAA AGTTTTATAA   | 1680 |
|    | ATCAAAATGAA AGGTGACATT GCCAGAACT TACCATTAACT TTGCACTACT AGGGTAGGGA  | 1740 |
| 50 | GGACTTAGGG ATGTTTCTCTG TGTCGTATGT GCTTTTCTTT CTTTCATATG ATCAATTCCTG | 1800 |
|    | TTGGTATTTT CAGTATCTCA TTTCTCAAAG CTAAAGAGAT ATACATTCTG GATACTTGGG   | 1860 |
| 55 | AGGGGAATAA ATTAAAGTTT TCACACTGNA AAAAAAAAAA AAAAAAAAC TCGA          | 1914 |

(2) INFORMATION FOR SEQ ID NO: 35:

60

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1020 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

5

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

CCNITNNITTT TTTTITTTTG CAAGACAAAA TATACTTTAT TGTGACAGCA AATGCACATA 60  
 10 GTGCTGTAGG TAAGGCATGC TACTAGGAAT CTGCATATAA TCAAAAGCCA GTATGGAAAT 120  
 GAATGGAAAT GAATGCTGTT GTTCTCAGAT TGAGTCCATG GTGGAGAAAG GATAGTTTGT 180  
 GTCCACTTAT TTCAAATGCA GTATCATACC TACTTAATCA GTTACCTATG CTTCTAACCA 240  
 15 ACAGCCCACT GGCAAATAGG AGGAACCTAA CTGTACTCAG AAGTCACITTT TAATATCAAC 300  
 GACAGAAATA TTTCACTAAT TCAACTGAGG CAAATTTCTT TTCTAGACAA AGGACCTAGA 360  
 20 AATTGAGCAT GCAAACATC CATCCATTCA TTCATTCAAA TAATTAGCCA ATTTTACCGT 420  
 CATTTAATTC CACCAGAAGC AAATACTAGA ATATCTAGAA GTAGTTTGGG TAAAGAAACA 480  
 25 TTTACATTTT AATATTGTGT AATGTCATAA ATTTGGGGCT AAAATAACAC CAGGTCAAAT 540  
 TTGATCCCTT TGTATGTGAG GGTACAAAGT ACAGTTTTCG TTTCAACAGC TGAACITCTG 600  
 AGAGAAGAGC TGAAAAAAT GCTAAATAAG AGATCTAGGC CTTTGATGGA AACTATTAGG 660  
 30 CTCTACAGAC TTGTCAAAAA ATCAATGCAA AACTGAGGGG GAAAGGCTGA AATGCTTTGT 720  
 AAAGCAGTAT TTTTAGACAA GTTGCTTCAT TTCCCCCTTT TCTAAAACAG ATGCAGATTA 780  
 AATGTTTTTT TGCAATGATG CACATTGACA TTCTGTTCAA CTGTTTTCTA AATGCAACAC 840  
 35 TGCGGGTTTC AACAGTATGC TTTCAATTA ACAAAGAATA TTATATGCAT GGTCAATTTA 900  
 GTTTAAGAGA TGAAAAAATA CTTTACTACT ATGAAAATTG CTTATCAAAT ACTCTCCTCT 960  
 40 TTTATAAGGT GTTTTARGC AACACAGGAC CGGTNGAACC GANCAAATTT ATAATTATAC 1020

45

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

50

- (A) LENGTH: 781 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

55 AACTCCTGAC CTCAAGTGCT CCACCTGCGT TGGCTTCCCA AAGTGCTGGG ATACAGGAGT 60  
 RAGCCACTGC GCCTGGCTGA TCCCAGCACT TTMAAAATGA TGCCGCTCAA AGCCGTGACT 120  
 60 TGGCCTACTT TGAACAGCAA ACTTGTGCT GCTGTGTGCA ACCTGAAGGC CTCTCAAATG 180

CCAGCTTCAA GCAGGGTGTG AATTGGCCAG TGTGAGATCT CAGGAGTCCT GTGTTGAGAG 240  
 TGTGGCTTTC AGCTGCGGGG AGCTGCACTT GGTGGGGAAA GCCAGGCAGG TCACCCTCAC 300  
 5 AGCCAGATAA TGTGGAGGTC AGAACCCAAG GAAGGGAGTG AGACCTCCAC TCCCAGTGGG 360  
 GGACCTGGCC ACCCATCCTT GGGGACCTGA GAAAGCGTAC TTCACCTTGG GGTGAAGGCT 420  
 10 GGGTGGGGCC AGAGGGACCA GTGCCCTCCT CAGTGCTTAG GGGCAGAGCC ACCTGCAGCA 480  
 ATGGTATCTG CATATTAGCC CCTCTCCACC TTCTTTCTCC CGCTGAATCA TTTCCCTCAA 540  
 AGCCCAAGAG CTGTCACTGC TTCTTTCTCC CTGGGAAGAA TGCCTGGACT CTGCCTGGTG 600  
 15 ATAGACTGAA GCCAGAACAG TGCCACACCC TCGCCTTAAT TCCTTGCTAG GTGTTCTCAG 660  
 ATTTATGAGA CTTCTTAGTC AAATATGAGG GAGGTGGAT GTGGTGGCTT GTGCCTGTAA 720  
 20 TCCCAGCATT TTGGGAAGCC GAGGTGGGAG GATCCCTTGA AGCCAGGAGT TTGAGACAAG 780  
 C 781

25

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 966 base pairs  
 30 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

35 GGCACGAGGA AGCAGCTGGG GGCTGATCAG GGGGAGCACG CAGCCCTCCG ATTGCAGGGC 60  
 TGCCTATTGT AGTGGCAGCT CCTCTTGAAA CAATGCAGAA CAAGCCCAGG GCCCCACAGA 120  
 40 AAAGGGCACT GCCCTTCCCA GAACTTGAGC TCOGGGACTA CGCATCTGTT CTCACCAGAT 180  
 ACAGCTTGGG GCTGAGGAAC AAAGAGCCTT CCCTGGGCCA CAGGTGGGGG ACCCAGAAGC 240  
 45 TGGGCAGGAG CCCCTGTCTT GAAGGGTCCC AGGGCCACAC CACAGATGCT GCTGACGTGC 300  
 AGAACCACTC TAAAGAAGAA CAGAGAGACG CAGGAGCACA GAGGARGTGC GGCCAGGGGA 360  
 GGCACACCTG GCGGTACAGG NGAGGGGCGC AGGACACTTC GAGGCTGACA GGAGACCCAC 420  
 50 GTGGTGGGGA AAGGAGCCCC CCAAAGTGTG AGAGCATGAA GCAGCAGGAA GGAGCTCCCT 480  
 CGGGCCACTG CTGGGATCAG TGGTGCCATG GAGCAAGCGA GGTGTGTTGG CCTGAAAGCC 540  
 GGAAGCGTGC CCAATCTTT SCATCACCAT GTAGGCAGTC ACCTCGCTCC TCAGCACTCG 600  
 55 GGCAGGACA GAAGCTTGCT GTCTGCTCAC CAGACATCCT GTGCTGCCCT ACAGACACCT 660  
 TGCTCGCCAG CCATCCCCAC TCACTTCTGA CCGGGACCCA ATTCTCTGGC CAAACCCAGG 720  
 60 CTCTAGCACC GTCTTGGTGT GCTTGAGAAA CATCTAGTTT AAGTCAAAT CCAATGTCTT 780

TTTAATATAT AGACTATATG TACCTATGGA CTAGAGGTGA ATATATATAC ATCATATCAA 840  
 5 ATTCAAGTGA CCCAGTATTT CGGGAGAACC CACTATGTTC CCAGCCTGCA TGGGAAGCTG 900  
 GGGATTCTGG CATGAACTGC ACCTTATCTT CCTCGAGGGG GGGCCGGTAC CAATTGCCNA 960  
 TAGTGG 966

10

(2) INFORMATION FOR SEQ ID NO: 38:

15

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 416 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GAATTCGGCA CGAGGTAATA GGAGCCCTCG TACCTCTTGT GTTCCTTACA AACATTCTCA 60  
 25 TCAGTAGCTC TACGCGTTGA CTGGGTGGTT TGARATGGCT GGTATACACA GGGCTTTCTT 120  
 GGTGTCTGT CTCTGGGGCT TACCTTTGTG TGTGGTTGGA GGCCCTGGT GAGATTGGAA 180  
 30 GTACCAGAGA GTGCTGTGTC AGGGGCAGAG GGGCCTGTG CTGGAGCTGG AGGTGCTCTG 240  
 CCTTGTGTC TGACTCARTC TCCTGTCTGC CTGCCCCCT CAGGGTCTCG CCAGCCGAGC 300  
 CTCTGTGGGA ATCTAAAAGG ARTGGATGTG GACGKTGAC CAAGCACATC TCAGCTTTTA 360  
 35 ATACCTGGGC TATTATAGA CCTTTGGGGG GAATNGCTTG TGAACAACA AGGGTT 416

40

(2) INFORMATION FOR SEQ ID NO: 39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1114 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

50 TGTGTATTTG GGGGACTGA AGGTACGTG GGGCGAAACA AAACCGGCCA TGGCAGCAGC 60  
 GGAGGAGGAG GACGGGGGCC CCGAAGCCAA AATCGCGAGC GGGCGGGGC GGGCGGACC 120  
 55 TTCGAATGTA ATATATGTTT GGAGACTGCT CGGGAAGCTG TGGTCAGTGT GTGTGGCCAC 180  
 CTGTACTGTT GGCCATGTCT TCATCAGTGG CTGGAGACAC GGCCAGAACG GCAAGAGTGT 240  
 CCAGTATGTA AAGCTGGGAT CAGCAGAGAG AAGGTTGTCC CGCTTTATGG GCGAGGGAGC 300  
 60 CAGAAGCCCC AGGATCCCAG ATTAAAACT CCACCCCGCC CCCAGGGCCA GAGACCAGCT 360

5 CCGGAGAGCA GAGGGGGATT CCAGCCATTT GGTGATACCG GGGGCTTCCA CTTCTCATTT 420  
 GGTGTGGTG CTTTCCCTT TGGCTTTTC ACCACCGTCT TCAATGCCCA TGAGCCTTTC 480  
 CGCCGGGGTA CAGGTGTGGA TCTGGGACAG GGTCACCCAG CCTCCAGCTG GCAGGATTCC 540  
 CTCTTCCTGT TTCTCGCCAT CTTCTTCTTT TTTTGGCTGC TCAGTATTTG AGCTATGTCT 600  
 10 GCTTCCTGCC CACCTCCAGC CAGAGAAGAA TCAGTATTGA GGTCCCTGC TGACCCCTCC 660  
 GTACTCCTGG ACCCCCTTGA CCCCTCTATT TCTGTTGGCT AAGGCCAGCC CTGGACATTG 720  
 TCCAGGAAGG CCTGGGGAGG AGGAGTGAAG TCTGTGCATA GATGGGAGAG CCTTCTGCTC 780  
 15 AGAGGCTCAC TCAGTAACGT TGTTTAATTC TCTGCCCTGG GGAAGGAGGA TGGATTGAGA 840  
 GAATGTCTTT CTCTCTCCT AAGTCTTTC TTTCCCTGAT TTCTTGATTT GATCTTCAAA 900  
 20 GGTGGGCAA GTTCCCTCTG ACTCTTCCCC CACTCCCCAT CTTACTGATT TAATTTAATT 960  
 TTTCACTCCC CAGAGTCTAA TATGGATTCT GACTCTTAAG TGCTTCGGCC CCTCACTAC 1020  
 CTCTTTAAT ACAAAATCAA TAAAAAGGT GAAATATAAA AAAAAAAAAA AAAAAACYCG 1080  
 25 GGGGGGGCCC CGGTCCCCAT TCCCTTTGGG GGGT 1114

30

(2) INFORMATION FOR SEQ ID NO: 40:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 602 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

GGGTCGACCC ACGCGTCCGT CCCAGGCCAC AAGACATTTT CTGCTCGGAA CCTTGTTTAC 60  
 TAATGTCTTC TGTGGCACAT TTTGTTTCCC GTGCCCTGGG TGTCAAGTTG CAGCTGATAT 120  
 45 GAATGAATGC TGTCTGTGTG GAACAAGCGT CGCAATGAGG ACTCTCTACA GGACCCGATA 180  
 TGGCATCCCT GGATCTATTT GTGATGACTA TATGGCAACT CTTTGCTGTC CTCATTGTAC 240  
 TCTTTGCCAA ATCAAGAGAG ATATCAACAG AAGGAGAGCC ATGCGTACTT TCTAAAAACT 300  
 50 GATGGTGAAA AGCTCTTACC GAAGCAACAA AATTCAGCAG ACACCTCTTC AGCTTGAGTT 360  
 CTTCAACATC TTTTGCAACT GAAATATGAT GGATATGCTT AAGTACAACT GATGGCATGA 420  
 55 AAAAAATCAA ATTTTGTATT TATTATAAAT GAATGTTGTC CCTGAACCTA GCTAAATGGT 480  
 GCAACTTAGT TTCTCCTTGC TTTTATATTA TCGAATTTCC TGGCTTATAA ACTTTTTTAA 540  
 60 TTACATTTGA AATATAAACC AAATGAAATA TTTTACTGAA AAAAAAAAAA AAAAAANCCC 600

CA

602

5

(2) INFORMATION FOR SEQ ID NO: 41:

(i) SEQUENCE CHARACTERISTICS:

10

- (A) LENGTH: 970 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

15

|  |     |
|--|-----|
| GGCAGAGCTT AGGAGAACAG CTCCTTTGG ATCCCTNTCA AAGGTGATAC CATGGGCTCC   | 60  |
| CAGCTTAGAG TAAGAAGCTC TGAGAAGTGT AATGAAGGGT GAGATAGAGA TGCTGAACCC  | 120 |
| ATTCTTSCAG CTCTCTCTAG TGTGTATTAT TCCAGAATGG CCAACACCCC TACATTGATA  | 180 |
| CATAAACACA TTCCAAGGCC TTGTGTAATA CAAAGTTCAC CGTCCTCTCTG GAATAGGAGC | 240 |
| CCTGGGTTCT AGTTCTCACT CTGCCACTGG GGGAAAATCC AATTAAAGTC TGGTTTAGTC  | 300 |
| AGCTTGGGTC ACCATAGACT GGGTGGCTTA AACAGCAGAC ATTTATTTCT GGTAGTTTCT  | 360 |
| GGAGCTACA AATCTAAGAG CAAGGTGCCA GCATGGTCAC ATTCTGGTGA GGGSCCTCTT   | 420 |
| CCTGGCTTGT AGACGGCTGC YTTCTCACCG TGTGCTCACA TAGCCTTTCTG TGTGTGTGTG | 480 |
| TGTGTGTGTG TCGCTKGTG CAAGCTTCK GATGTCTCTT CTTAGAAGGA CACCAACCCC    | 540 |
| ATCATGAGAG CCTACTCTC ATGACTTAGC CTAACCCTAA TTACCCTCCA AAGGCCCCAT   | 600 |
| CTCCAAATGC CATCACATTG GAGGGTAGAG CTTCAACATA GGGATTTTGG GGGACACAAA  | 660 |
| CATTCAGTCC ATAACAAAGG CTGTAGTCTT TARTTCTCT GTCTGTGAAA TGAGAGTGT    | 720 |
| GAGATTCTTT CTAGCCTTTA TCATTTATAA TTCTGTGAGA TGTAGATTG CATTATTTTC   | 780 |
| GAGTTCGAGT TATATGAAAT GTTCCCTCT ACATTTTCTT GGGCAACTGA GAACTGAATA   | 840 |
| GGGCTAGGTT TAAATAGAGT TAGGCAGTTA GGCTTATTCT TTTATTTAAT AAGCATTTTT  | 900 |
| GGAGCATCTA CGGTGTTCCA GGAAGTGAAC TGTGTAAAC ATTGGAGCTG TAACAGAGAA   | 960 |
| CAAAAGAGAC   | 970 |

50

(2) INFORMATION FOR SEQ ID NO: 42:

55

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1002 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

60

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

|    |  |      |
|----|--|------|
|    | GAATTCGGCA CGAGCCGAGG TCGGCAGCAC AGAGCTCTGG AGATGAAGAC CCTGTTCTCTG | 60   |
| 5  | GGTGTACGCG TCGGCTGGC CGCTGCCCTG TCCTTMACCC TGGRGGAGGA GGATATCACA   | 120  |
|    | GGGACCTGGT ACGTGAAGGC CATGGTGGTC GATAAGACTT TCCGGAGACA GGAGGCCCAG  | 180  |
| 10 | AAGGTGTCCC CAGTGAAGGT GACAGCCCTG GCGCGTGGGA AGTTGGAAGC CACGTTACCC  | 240  |
|    | TTCATGAGGG AGGATCGGTG CATCCAGAAG AAAATCCTGR TCGGAAGAC GGAGGAGCCT   | 300  |
|    | GGCAAATACA GCGCTGTGA GCGCTCCCC CAYTCCCACC CCCACCTCC CCCACCGCCA     | 360  |
| 15 | ACCCCACTGC ACCAGCTCC ACAGTAGAG AGTGCCAGG CTGCCCTTTT GCCAGGGCCC     | 420  |
|    | CAGCTCTGCC CACCTCCAAG GAGGGGCTGG CCTCTCTTC CTGGGGGCT GGTGGCCCTG    | 480  |
| 20 | ACATCAGACA CCGGTGTGA CAGGCTGTG CGCAGTCGAG ATGGACCAGA TCACGCCTGC    | 540  |
|    | CCTCTGGGAG GCGCTAGCCA TTGACACATT GAGGAAGCTG AGGATTGGGA CAAGGAGGCC  | 600  |
|    | AAGGATTAGA TGGGGGAGG AAGCTCATGT ACCTGCAGGA GCTGCCAGG AGGGACCAAT    | 660  |
| 25 | ACATCTTTTA CTGCAAGAC CAGCACCATG GGGGCTGCT CCACATGGGA AAGCTTGTGG    | 720  |
|    | GTAGGAATTC TGATACCAAC CCGGAGGCC TGAAGAATT TAAGAAATTG GTGCAGCGCA    | 780  |
| 30 | AGGGACTCTC GGAGGAGGAC ATTTTCACGC CCTGTCAGAC GGAAGCTGC GTTCCCGAAC   | 840  |
|    | ACTAGGCAGC CCGGGGTCT GCACCTCCAG AGCCACCCCT ACCACCAGAC ACAGAGCCCG   | 900  |
|    | GACCACCTGG ACCTACCTC CAGCCATGAC CCTTCCCTGC TCCACCCAC CTGACTCCAA    | 960  |
| 35 | ATAAAGTCCT TCTCCCCCA AAAAAAAAAA AAAAAAACTC GA                      | 1002 |

## 40 (2) INFORMATION FOR SEQ ID NO: 43:

## (i) SEQUENCE CHARACTERISTICS:

- 45 (A) LENGTH: 2581 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

|    |   |     |
|----|---|-----|
| 50 | TGCAAAACCA CTGGACACTG GACAAGTACG GGATCCTGGS CGACGCACGC CTCTTCTTTG | 60  |
|    | GGCCCCAGCA CCGGCCGTC ATCCTTCGGT TGTCCAACCG CCGCGCACTG CGCCTCCGTG  | 120 |
| 55 | CCAGCTTCTC CCAGCCCCTC TTCCAGGCTG TGGSTGCCAT CTGCCGCTC CTCAGCATCC  | 180 |
|    | GGCACCCCGA GGAGCTGTCC CTGCTCCGG CTCTGAGAA GAAGGAGAAG AAGAAGAAAG   | 240 |
|    | AGAAGGAGCC AGAGGAAGAG CTCTATGACT TGAGCAAGGT TGTCTTGGCT GGGGGCGTGG | 300 |
| 60 | CACCTGCACT GTTCCGGGGG ATGCCAGCTC ACTTCTCGGA CAGCGCCCAG ACTGAGGCCT | 360 |

|    |   |      |
|----|---|------|
|    | GCTACCACAT GCTGAGCCGG CCCAGCCGC CACCCGACCC CCTCCTGCTC CAGCGTCTGC  | 420  |
| 5  | CACGGCCCAG CTCCTTGTC GACAAGACCC AGCTCCACAG CAGGTGGCTG GACTCGTCGC  | 480  |
|    | GGTGTCTCAT GCAGCAGGGC ATCAAGGCCG GGGACGCACT CTGGCTGCGC TTCAAGTACT | 540  |
|    | ACAGCTTCTT CGATTTGAT CCCAAGACAG ACCCCGTGCG GCTGACACAG CTGTATGAGC  | 600  |
| 10 | AGGCCCCGTG GGACCTGCTG CTGGAGGAGA TTGACTGCAC CGAGGAGGAG ATGATGGTGT | 660  |
|    | TTGCCGCCCT GCAGTACCAC ATCAACAAGC TGTCCCAGAG CGGGGAGGTG GGGGAGCCGG | 720  |
| 15 | CTGGCACAGA CCCAGGCTG GACGACCTGG ATGTGGCCCT GAGCAACCTG GAGGTGAAGC  | 780  |
|    | TGGAGGGGTC GGGCCCCACA GATGTGCTGG ACAGCCTCAC CACCATCCCA GAGCTCAAGG | 840  |
|    | ACCATCTCCG AATCTTTCGG CCCCGGAAGC TGACCCTGAA GGGCTACCGC CAACACTGGG | 900  |
| 20 | TGGTGTTCAA GGAGACCACA CTGTCTACT ACAAGAGCCA GGACGAGGCC CCTGGGGACC  | 960  |
|    | CCATTCAGCA GCTCAACCTC AAGGGCTGTG AGGTGGTTCC CGATGTTAAC GTCTCCGGCC | 1020 |
| 25 | AGAAGTTCTG CATTAACTC CTAGTGCCCT CCCCTGAGGC ATGAGTGAGA TCTACCTGCG  | 1080 |
|    | GTGCCAGGAT GAGCAGCAGT ATGCCCGCTG GATGGCTGGC TGCCGCTGG CCTCCAAAGG  | 1140 |
|    | CCGCACCATG GCCGACAGCA GCTACACCAG CGAGGTGCAG GCCATCCTGG CYTTCCTCAG | 1200 |
| 30 | CCTGCAGCGC ACGGGCAGTG GGGGCCCGGG CAACCACCCC CACGGCCCTG ATGCCTCTGC | 1260 |
|    | CGAGGGCCTC AACCCTTACG GCCTCGTTGC CCCCCGTTTC CAGCGAAAGT TCAAGGCCAA | 1320 |
| 35 | GCAGCTCACC CCACGGATCC TGAAGGCCA CCAGAATGTG GCCCAGTTGT CGCTGGCAGA  | 1380 |
|    | GGCCCAGCTG CGCTTCATCC AGGCCTGGCA GTCCCTGCCC GACTTCGGCA TCTCCTATGT | 1440 |
|    | CATGGTCAGG TTCAAGGGCA GCAGGAAAGA CGAGATCCTG GGCATCGCCA ACAACCGACT | 1500 |
| 40 | GATCCGCATC GACTTGGCCG TGGGCGACGT GGTCAAGACC TGGCGTTTCA GCAACATGCG | 1560 |
|    | CCAGTGGAAT GTCAACTGGG ACATCCGGCA NGTGGCCATC GAGTTTGATG AACACATCAA | 1620 |
| 45 | TGTGGCCTTC AGCTGCGTGT CTGCCAGCTG CCGAATTGTA CACGAGTATA TCGGGGGCTA | 1680 |
|    | CATTTTCCTG TCGACGGGG AGNGGGCCCG TGGGGAGGAG CTGGATGAAG ACCTCTTCCT  | 1740 |
|    | GCAGCTCACC GGGGGCCATG AGGCCTTCTG AGGGCTGTCT GATTGCCCTT GCCCTGCTCA | 1800 |
| 50 | CCACCTGTC ACAGCCACTC CCAAGCCCAC ACCCACAGGG GCTCACTGCC CCACACCCGC  | 1860 |
|    | TCCAGGCAGG CACCCAGCTG GGCATTTTAC CTGCTGTAC TGACTTTGTG CAGGCCAAGG  | 1920 |
| 55 | ACCTGGCAGG GCCAGACGCT GTACCATCAC CCAGGCCAGG GATGGGGGTG GGGGTCCCTG | 1980 |
|    | AGCTCATGTG GTGCCCCCTT TCCTTGTCTG AGTGGCTGAG GCTGATACCC CTGACCTATC | 2040 |
|    | TGCAGTCCCC CAGCACACAA GGAAGACCAG ATGTAGCTAC AGGATGATGA AACATGGTTT | 2100 |
| 60 | CAAACGAGTT CTTTCTGTT ACTTTTAAA ATTTCTTTT TATAAATTAA TATTTTATG     | 2160 |

TTGGATCCTC CTCCTTTCTC TGGAGCTGTG CTGGGGGCTA CTCTGACACT CTGTCTCTTC 2220  
 5 ATCACCAGCC AAGGAAAGGG GCTTTCCTGA TAAAGACAAG AGTTGGTTAG AGAAAGGGAC 2280  
 ACCTAAGTCA GTCTAGGGTT GGAAGCTAGG AGAGAGGTGA GGCAGAAGG GCACAGCTTT 2340  
 CAGGAACAAG GAATAGGGGC TGGGGTKGTK GTTCTCACGG GTAGGCGGTA CCTGCAGGGC 2400  
 10 CTCTTTGAAG TACTTGGGAA GGAGGAAGCC ATCAGTATTC CCTGGAGTCA GAATCACCCC 2460  
 ATTGGCAGAG CGAAGAAGG GTATTCATC TGCTGACAGA GCCAGAGATG TGAATCATGC 2520  
 15 CCTCCCGAA GGCAAAGTCA GCTCTGCTT TGTCCAGACT CACCTGCCAG AGCCAGGGGT 2580  
 C 2581

20

(2) INFORMATION FOR SEQ ID NO: 44:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1764 base pairs  
 25 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

30 GAATTCGGCA CGAGGATGAT ATTCTACTA TTCCTCACC CACTCTGGCT GCAAAAAGGA 60  
 AGTGCAGGGA AAATGAGTGG GGAGTTCCTG TATGCCAGTC TGTTCATG GAACTATTTT 120  
 35 TGGAGGAATA AAAAGTATG CTAGATTATA TTGGTACGAT AGGCATTTTC TTACATGCA 180  
 TATAGTCTGC TTTGGCTTTT ACCTGTGAG GGAAGAATG AGGAGAGGAT AAAAATCATT 240  
 GTATCCCTA GAGAAGGAAT ATCAAATCC ATTTAATAAA AAACTCATA CTAAGAATAA 300  
 40 AATTGCATAG TGTTTTATTC TCCTTTGTTC ATAATTAAAC ACAAGATATT TTAAATGTG 360  
 AAATCAGTTT CTTTATGAAA AAATATGACC TGTATGCCTT TATTCTCTCC TTTCTCTCTT 420  
 45 CCCACCCGTC GCTTCTTTTC TTCTCTTCCT TTTTTCCTT CCTTGTCTC TGAATAATG 480  
 AAGAACAAAC ATTTGATAAA AGCCACTGCC AATTCATGAT AAAAATTCAC AGCAAAGTTG 540  
 GTACAGAAAA GAACTTTCTC TCGGTGTTAA AGGGTGCTC TCCATGCTC TCAGCAAATA 600  
 50 TTAAATGATG AATCTTATT AATAATCACT GTAGAACCA GAATTAACT AGTATACCCA 660  
 CTGTCTGGC TTGTAATCAA CAATATACAG GTGGTTCTAG CCAGTGCAAT AAGACAAGAG 720  
 55 AAACAAAAAT GTTATAAGGC CTGGAAGA GAACAAAC TGTATTAC AAAATACTGT 780  
 CTATACAGAA TGCTCAGTG CTTTCTTCTT TTCTTTTCTT TAAACTTTA GTGAGATACC 840  
 60 CTCTGCCCT ATCTTAAAT CACGTGGTGG GGGGTGGTGT CTGCACTGA AACAGGACAC 900

|    |   |      |
|----|---|------|
|    | TTGGTTCCTG GGTTTAGCAT TGACCTTGCC AGCTTGGTYT GGCAGCTGAG TTGTTGGACT | 960  |
|    | AGGAAGCGTC CYTGCAGGTT GTGKTCTGKT ACCTCTCTGT AAAGCCTGAA AGCATCCTAC | 1020 |
| 5  | SATTGCATTT GCTAGKTCTC AGTAGAGCTA TTTAACAAGA ATCTGGAAAC ATTTTYCCTG | 1080 |
|    | AGGGCTCTCT TTAGACAGCA GTAAATGTA GCTGGAGACA TATTGAGTAA ATGGAAGA    | 1140 |
| 10 | AAAATCTAAT GAGGCCAGGA ATTTTTTTAA TCTTCTATT TCACAGAAG CCTCAAGGAG   | 1200 |
|    | AACACCATAA TTCATATTTT ACTCAKGTGG GTTAGGCATA AAGCCTCCCC CATAGATCCA | 1260 |
|    | ATAACCTGTA RGTGTYCTGG TTTTGAAATT GCACTGCTT ACATKGCTGG ATCNTAGCAC  | 1320 |
| 15 | TAAWTCACAC RGCAACGGCT TCTGGTTCAA TKGTTCAITA CTTGGGAATG TCAGATTGCC | 1380 |
|    | AGAGAGCAGC CTGATGTTTA CATCCAATCG GCAATGCCTT AGGAAATCAG TTTTAATTAC | 1440 |
| 20 | AATCTCACGT AGCAGCACTG CACTCAACCT TCAGAGAGGC TGGGATTTGT GTTGAACCTA | 1500 |
|    | CATCTTATAG CTGTGCAGAA AATGCCTGTC CGACTGGGTC ATGCAAAATG GACAGCAAAG | 1560 |
|    | TCAGCAGAAC CTTAGAAAAG ATGACACAGC AAGTGAACA CAGCTGGATC ATCCCCCGTC  | 1620 |
| 25 | CTGTCAAGCG TGCAGTGCTC TCTGGCCCTT TTTTAAACA AGGGAACCCA GTTGGCGTTT  | 1680 |
|    | GCCTTTCAGC TTCCCCATTC TGATATAAAA ATCTGTGACC CAGCAGCTTT AACCATAAAA | 1740 |
| 30 | AAAAAAAAA AAAAAAAAAAC TCGA  | 1764 |

35 (2) INFORMATION FOR SEQ ID NO: 45:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 796 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: double
  - (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

|    |  |     |
|----|--|-----|
| 45 | ACCTTCTTCC ATGTTTAGTC CCTTGGGCTC TGCTACCCCTC CTGCTGGAGG TGAGAGCATC | 60  |
|    | CTGTGTGCAA CCAGAGATGC CCTCTGGCTT TCAGACCTGC CTGCTTTTCA CCCTCAGCCC  | 120 |
|    | TTTCTCACTC AGCAAAATG TGGGGTCCC TAGTCAGCAG CTCCCTGGGC AGCTCTCTGA    | 180 |
| 50 | GCAAGGTGGT CTCTGTGTC ATGAAGGAGA GCCGGCTAGG ACAGTGCCGG AAACCTCAGCT  | 240 |
|    | GCCTCTCCCC TTCAACTCAG CTGGCCCCC GCACCTGAAG TGCACAGGAG CCGGAAGAG    | 300 |
| 55 | AGTCTGGAGC CCACCCCGGA GGGCAGCACA GGAGGTGTCT YTGAGCTGG TGTCTGCM     | 360 |
|    | CCCYTGCAGG CAGMACAGT CCCGGGCATT YTCYTAGCC ACAGACAGAA CAGCCAGTGC    | 420 |
|    | CAGAGTCTGC TGTGTYTCC CCTTTAAGCA CACTCATTCA CCACACCCGA GGAGGCCAGA   | 480 |
| 60 | GGTGCAGGGA GCATGGGCTG TCGTTCCTT TTAAGCACAC TCATTACCA CACCCGAGGA    | 540 |

5 GGCCAGAAGT GCAGGGAGCA TGGGCTGGGT GCACCTCCGC AGGAGAGAAG GCTGAGCCAC 600  
 CGCCGTCCCG GGAGCCCGGC TCCCAGGCCT CTCGTTTTC CCTACCTCCC TAAGACTTTT 660  
 CTGTCACTCT CTGGCCATTG AAAGGCTTCT GTTCCTTAAA GTGCTGTAC ACTCTCCTTT 720  
 CCCAGGATGC AGCAAGCCAA AACAGTACCA CTGCACGTCA GCCTGGGTGA CAGAGTGAGA 780  
 10 CCTATCTTA AAAAAA 796

15 (2) INFORMATION FOR SEQ ID NO: 46:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1705 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

25 TGGCCATGGA AGCGCTAGAA GGTTTAGATT TTGAAACAGC AAAGAAGGAT TTCCTTGGAT 60  
 CTGGAGACCC CAAAGAAACA AAGATGCTAA TCACCAAACA GGCTGACTGG GCCAGAAATA 120  
 TCAAGGAGCC CAAAGCOGCC GTGGAGATGT ACATCTCAGC AGGAGAGCAC GTCAAGGCCA 180  
 30 TCGAGATCTG TGGTGACCAT GGCTGGGTTG ACATGTTGAT CGACATCGCC CGCAAACCTGG 240  
 ACAAGGCTGA GCGCGAGCCC CTGCTGCTGT GCGCTACCTA CCTCAAGAAG CTGGACAGCC 300  
 35 CTGGCTATGC TGCTGAGACC TACCTGAAGA TGGGTGACCT CAAGTCCCTG GTGCAGCTGC 360  
 AGTGGAGACC CAGCGCTGGG ATGAGGCCTT TGCTTTGGGT GAGAAGCATC CTGAGTTTAA 420  
 GGATGACATC TACATGCCGT ATGCTCAGTG GCTAGCAGAG AACGATCGCT TTGAGGAAGC 480  
 40 CCAGAAAGCG TTCCACAAGG CTGGGCGACA GAGAGAAGCG GTCCAGGTGC TGGAGCAGCT 540  
 CACAAACAAT GCCGTGGCGG AGAGCAGGTT TAATGATGCT GCCTATTATT ACTGGATGCT 600  
 45 GTCCATGCAG TGCTCGATA TAGCTCAAGA TCCTGCCCAG AAGGACACAA TGCTTGGCAA 660  
 GTTCTACCAC TTCCAGCGTT TGGCAGAGCT GTACCATGGT TACCATGCCA TCCATCGCCA 720  
 CACGGAAGAT CCGTTCAGTG TCCATCGTCC TGAAACTCTT TTCAACATCT CCAGTTTCTT 780  
 50 GCTGCACAGC CTGCCCAAGG ACACCCCTC GGGCATCTCT AAAGTGAAAA TACTCTTCAC 840  
 CTTGGCCAAG CAGAGCAAGG CCCTCGGTGC CTACAGGCTG GCCCGGCACG CCTATGACAA 900  
 55 GCTCGTGGC CTGTACATCC CTGCCAGATT CCAAAGTCC ATTGAGCTGG GTACCCTGAC 960  
 CATCCGCGCC AAGCCCTTCC ACGACAGTGA GGAGTTGGTG CCCTTGTGCT ACCGCTGCTC 1020  
 60 CACCAACAAC CCGCTGCTCA ACAACCTGGG CAACGTCTGC ATCAACTGCC GCCAGCCCTT 1080

CATCTTCTCC GCCTCTTCCT ACGACGTGCT ACACCTGGTT GAGTTCTACC TGGAGGAAGG 1140  
 GATCACTGAT GAAGAAGCCA TCTCCCTCAT CGACCTGGAG GTGCTGAGAC CCAAGCGGGA 1200  
 5 TGACAGACAG CTAGAGATTT GCAAACAACA GCTCCCAGAT TCTTGCGGCT AGTGGGAGAC 1260  
 CAAGGGACTC CATCGGAGAT NAGGACCCGT TCACAGCTAA GCTRAGCTTT GAGCAAGGTG 1320  
 10 GCTCARAGTT CGTGCCAGTG GTGGTGAGCC GGCTGGTGCT GCGCTCCATG AGCCGCCGGG 1380  
 ATGTCTCAT CAAGCGATGG CCCCCACCCC TGAGGTGGCA ATACTTCCGC TCACTGCTGC 1440  
 CTGACGCCTC CATTACCATG TGCCCCCTCT GCTTCCAGAT GTTCCATTCT GAGGACTATG 1500  
 15 AGTTGCTGGT GCTTCAGCAT GGCTGCTGCC CCTACTGCCG CAGGTGCAAG GATGACCCTG 1560  
 GCCCATGACC AGCATCTTGG GGACGGCTG CACCTCTGC CCGCCTTGGG GTCTGCTGGG 1620  
 20 CTGTGAAGGA GAATAAAGAG TTAAACTGTC AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1680  
 AAAAAAAAAA AAAAAAAAAA AAANA 1705

25

(2) INFORMATION FOR SEQ ID NO: 47:

(i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 981 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

35 TCGGCAGCAC AGAGCTCTGG AGATGAAGAC CCTGTTCTTG GGTGTACGC TCGGCCTGGC 60  
 GCTGCCCTGT CCTTCACCCT GGRGGAGGAG GATATCACAG GGACCTGGTA CGTGAAGGCC 120  
 40 ATGGTGGTCG ATAAGACTTT COGGAGACAG GAGGCCAGA AGGTGTCCC AGTGAAGGTG 180  
 ACAGCCCTGG GCGGTGGGAA GTTGAAGCC ACGTTCACCT TCATGAGGGA GGATCGGTGC 240  
 ATCCAGAAGA AAATCCTGRT GCGGAAGACG GAGGAGCCTG GCAAATACAG CGCCTGTGAG 300  
 45 CCCCTCCCCC AYTCCACCC CCACCTCCC CCACCGCAA CCCAGTGCA CCAGCCTCCA 360  
 CAGGTAGAGA GTGCCCAGGC TGCCCTTTTG CCAGGGCCCC AGCTCTGCC ACCTCCAAGG 420  
 50 AGGGGCTGGC CTCTCCTTCC TGGGGGGCTG GTGGCCCTGA CATCAGACAC CGGTGTGAC 480  
 AGGCTTGTC GCAGTCGAGA TGGACCAGAT CACGCTGCC CTCTGGGAGG CCTAGCCAT 540  
 TGACACATTG AGGAAGCTGA GGATTGGGAC AAGGAGGCCA AGGATTAGAT GGGGCAGGA 600  
 55 AGCTCATGTA CCTGCAGGAG CTGCCCAGGA GGGACCAYTA CATCTTTTAC TGCAAAGACC 660  
 AGCACCATGG GGCSTGCTC CACATGGGAA AGCTTGTTGG TAGGAATTCT GATACCAACC 720  
 60 GGGAGGCCCT GGAAGAATTT AAGAAATTGG TGCAGCGCAA GGGACTCTCG GAGGAGGACA 780

TTTTCACGCC CCTGCAGACG GGAAGCTGCR TTCCCGAACA CTAGGCAGCC CCCGGGTCTG 840  
 CACCTCCAGA GCCCACCCTA CCACCAGACA CAGAGCCCGG ACCACCTGGA CCTACCCTCC 900  
 5 AGCCATGACC CTTCCTTGCT CCCACCCACC TGA CTCCAAA TAAAGTCCTT CTCCCCCAAA 960  
 AAAAAAAAAA AAAAAACTCG A 981

10

(2) INFORMATION FOR SEQ ID NO: 48:

15

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 146 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

20

Met His Tyr Gln Met Ser Val Thr Leu Lys Tyr Glu Ile Lys Lys Leu  
 1 5 10 15

25

Ile Tyr Val His Leu Val Ile Trp Leu Leu Val Ala Lys Met Ser  
 20 25 30

Val Gly His Leu Arg Leu Leu Ser His Asp Gln Val Ala Met Pro Tyr  
 35 40 45

30

Gln Trp Glu Tyr Pro Tyr Leu Leu Ser Ile Leu Pro Ser Leu Leu Gly  
 50 55 60

35

Leu Leu Ser Phe Pro Arg Asn Asn Ile Ser Tyr Leu Val Leu Ser Met  
 65 70 75 80

Ile Ser Met Gly Leu Phe Ser Ile Ala Pro Leu Ile Tyr Gly Ser Met  
 85 90 95

40

Glu Met Phe Pro Ala Ala Gln Pro Ser Thr Ala Met Ala Arg Pro Thr  
 100 105 110

Val Ser Ser Leu Val Phe Leu Pro Phe Pro Ser Cys Thr Trp Cys Trp  
 115 120 125

45

Cys Trp Gln Cys Lys Cys Met Pro Gly Ser Cys Thr Thr Ala Arg Ser  
 130 135 140

Ser Xaa

50

145

(2) INFORMATION FOR SEQ ID NO: 49:

55

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 312 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

60

Met Asn Ser Val Val Ser Leu Leu Leu Ile Leu Glu Pro Asp Lys Gln  
 1 5 10 15  
 5 Glu Ala Leu Ile Glu Ser Leu Cys Glu Lys Leu Val Lys Phe Arg Glu  
 20 25 30  
 Gly Glu Arg Pro Ser Leu Arg Leu Gln Leu Leu Ser Asn Leu Phe His  
 35 40 45  
 10 Gly Met Asp Lys Asn Thr Pro Val Arg Tyr Thr Val Tyr Cys Ser Leu  
 50 55 60  
 Ile Lys Val Ala Ala Ser Cys Gly Ala Ile Gln Tyr Ile Pro Thr Glu  
 65 70 75 80  
 15 Leu Asp Gln Val Arg Lys Trp Ile Ser Asp Trp Asn Leu Thr Thr Glu  
 85 90 95  
 Lys Lys His Thr Leu Leu Arg Leu Leu Tyr Glu Ala Leu Val Asp Cys  
 100 105 110  
 Lys Lys Ser Asp Ala Ala Ser Lys Val Met Val Glu Leu Leu Gly Ser  
 115 120 125  
 25 Tyr Thr Glu Asp Asn Ala Ser Gln Ala Arg Val Asp Ala His Arg Cys  
 130 135 140  
 Ile Val Arg Ala Leu Lys Asp Pro Asn Ala Phe Leu Phe Asp His Leu  
 145 150 155 160  
 30 Leu Thr Leu Lys Pro Val Lys Phe Leu Glu Gly Glu Leu Ile His Asp  
 165 170 175  
 Leu Leu Thr Ile Phe Val Ser Ala Lys Leu Ala Ser Tyr Val Lys Phe  
 180 185 190  
 Tyr Gln Asn Asn Lys Asp Phe Ile Asp Ser Leu Gly Leu Leu His Glu  
 195 200 205  
 40 Gln Asn Met Ala Lys Met Arg Leu Leu Thr Phe Met Gly Met Ala Val  
 210 215 220  
 Glu Asn Lys Glu Ile Ser Phe Asp Thr Met Gln Gln Glu Leu Gln Ile  
 225 230 235 240  
 45 Gly Ala Asp Asp Val Glu Ala Phe Val Ile Asp Ala Val Arg Thr Lys  
 245 250 255  
 Met Val Tyr Cys Lys Ile Asp Gln Thr Gln Arg Lys Val Val Val Ser  
 260 265 270  
 50 His Ser Thr His Arg Thr Phe Gly Lys Gln Gln Trp Gln Gln Leu Tyr  
 275 280 285  
 55 Asp Thr Leu Asn Ala Trp Lys Gln Asn Leu Asn Lys Val Lys Asn Ser  
 290 295 300  
 Leu Leu Ser Leu Ser Asp Thr Xaa  
 305 310  
 60

## (2) INFORMATION FOR SEQ ID NO: 50:

5

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

10

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

Gly Gly Cys Pro Arg Arg Arg Leu Val Leu Tyr Cys Leu Phe Gly Ser  
 1 5 10 15

15

Ala Gly Gly Gly Arg Ile His Ser Glu Ala Trp Phe Pro Lys Ala Trp  
 20 25 30

Pro Glu Ala Glu Lys Trp Leu Phe Ala Glu Leu Leu Arg Gly Xaa  
 35 40 45

20

## (2) INFORMATION FOR SEQ ID NO: 51:

25

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 467 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

30

Met Leu Ser Arg Pro Gln Pro Pro Pro Asp Pro Leu Leu Leu Gln Arg  
 1 5 10 15

35

Leu Pro Arg Pro Ser Ser Leu Ser Asp Lys Thr Gln Leu His Ser Arg  
 20 25 30

Trp Leu Asp Ser Ser Arg Cys Leu Met Gln Gln Gly Ile Lys Ala Gly  
 35 40 45

40

Asp Ala Leu Trp Leu Arg Phe Lys Tyr Tyr Ser Phe Phe Asp Leu Asp  
 50 55 60

Pro Lys Thr Asp Pro Val Arg Leu Thr Gln Leu Tyr Glu Gln Ala Arg  
 65 70 75 80

45

Trp Asp Leu Leu Leu Glu Glu Ile Asp Cys Thr Glu Glu Glu Met Met  
 85 90 95

50

Val Phe Ala Ala Leu Gln Tyr His Ile Asn Lys Leu Ser Gln Ser Gly  
 100 105 110

Glu Val Gly Glu Pro Ala Gly Thr Asp Pro Gly Leu Asp Asp Leu Asp  
 115 120 125

55

Val Ala Leu Ser Asn Leu Glu Val Lys Leu Glu Gly Ser Ala Pro Thr  
 130 135 140

Asp Val Leu Asp Ser Leu Thr Thr Ile Pro Glu Leu Lys Asp His Leu  
 145 150 155 160

60

Arg Ile Phe Arg Pro Arg Lys Leu Thr Leu Lys Gly Tyr Arg Gln His

|    | 165   | 170 | 175     |
|----|---|-----|---------|
|    | Trp Val Val Phe Lys Glu Thr Thr Leu Ser Tyr Tyr Lys Ser Gln Asp |     |         |
|    | 180   | 185 | 190     |
| 5  | Glu Ala Pro Gly Asp Pro Ile Gln Gln Leu Asn Leu Lys Gly Cys Glu |     |         |
|    | 195   | 200 | 205     |
| 10 | Val Val Pro Asp Val Asn Val Ser Gly Gln Lys Phe Cys Ile Lys Leu |     |         |
|    | 210   | 215 | 220     |
|    | Leu Val Pro Ser Pro Glu Gly Met Ser Glu Ile Tyr Leu Arg Cys Gln |     |         |
|    | 225   | 230 | 235 240 |
| 15 | Asp Glu Gln Gln Tyr Ala Arg Trp Met Ala Gly Cys Arg Leu Ala Ser |     |         |
|    | 245   | 250 | 255     |
|    | Lys Gly Arg Thr Met Ala Asp Ser Ser Tyr Thr Ser Glu Val Gln Ala |     |         |
| 20 | 260   | 265 | 270     |
|    | Ile Leu Ala Phe Leu Ser Leu Gln Arg Thr Gly Ser Gly Gly Pro Gly |     |         |
|    | 275   | 280 | 285     |
| 25 | Asn His Pro His Gly Pro Asp Ala Ser Ala Glu Gly Leu Asn Pro Tyr |     |         |
|    | 290   | 295 | 300     |
|    | Gly Leu Val Ala Pro Arg Phe Gln Arg Lys Phe Lys Ala Lys Gln Leu |     |         |
|    | 305   | 310 | 315 320 |
| 30 | Thr Pro Arg Ile Leu Glu Ala His Gln Asn Val Ala Gln Leu Ser Leu |     |         |
|    | 325   | 330 | 335     |
|    | Ala Glu Ala Gln Leu Arg Phe Ile Gln Ala Trp Gln Ser Leu Pro Asp |     |         |
| 35 | 340   | 345 | 350     |
|    | Phe Gly Ile Ser Tyr Val Met Val Arg Phe Lys Gly Ser Arg Lys Asp |     |         |
|    | 355   | 360 | 365     |
| 40 | Glu Ile Leu Gly Ile Ala Asn Asn Arg Leu Ile Arg Ile Asp Leu Ala |     |         |
|    | 370   | 375 | 380     |
|    | Val Gly Asp Val Val Lys Thr Trp Arg Phe Ser Asn Met Arg Gln Trp |     |         |
|    | 385   | 390 | 395 400 |
| 45 | Asn Val Asn Trp Asp Ile Arg Gln Val Ala Ile Glu Phe Asp Glu His |     |         |
|    | 405   | 410 | 415     |
|    | Ile Asn Val Ala Phe Ser Cys Val Ser Ala Ser Cys Arg Ile Val His |     |         |
| 50 | 420   | 425 | 430     |
|    | Glu Tyr Ile Gly Gly Tyr Ile Phe Leu Ser Thr Arg Glu Arg Ala Arg |     |         |
|    | 435   | 440 | 445     |
| 55 | Gly Glu Glu Leu Asp Glu Asp Leu Phe Leu Gln Leu Thr Gly Gly His |     |         |
|    | 450   | 455 | 460     |
|    | Glu Ala Phe   |     |         |
|    | 465   |     |         |
| 60 |   |     |         |

## (2) INFORMATION FOR SEQ ID NO: 52:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 83 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

5  
 10 Met Arg Pro Gly Arg Gly Ala Gly Thr Pro Gly Arg Pro Gly Arg Gly  
     1                    5                    10                    15  
     Arg Gly Leu Ala Ala Thr Cys Ser Leu Ser Ser Pro Ser His Leu Leu  
                     20                    25                    30  
 15 Pro Thr Leu Leu His Thr Phe Ser Phe Ser Leu Pro Pro Pro Ser Pro  
             35                    40                    45  
 20 Ala Ala Pro Arg Gln Pro Ser Pro Pro Ala Leu Leu Leu Pro Gly Pro  
         50                    55                    60  
     Gln Lys Pro Arg Pro Gly Asp Pro Thr Tyr Thr Gly Ala Leu Thr Asp  
         65                    70                    75                    80  
 25 Trp Ser Xaa

## 30 (2) INFORMATION FOR SEQ ID NO: 53:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 63 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

35  
 40 Met Phe Leu Val Phe Phe Leu Ser Phe Phe Ser His Ser Ile Ser Ala  
     1                    5                    10                    15  
     Leu Thr Leu Val Cys Ser Gln Gly Gly Lys Ala Asp Met Asn Leu Leu  
             20                    25                    30  
 45 Ser Trp Asp Phe Arg Pro His Trp Leu Glu Gly Ile Arg Phe Leu Leu  
         35                    40                    45  
     Gly Trp Gly Gln Ala Leu Met Ala Gly Leu Phe Pro Trp Leu Xaa  
         50                    55                    60

50

## (2) INFORMATION FOR SEQ ID NO: 54:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 124 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

55  
 60 Met Arg Gly Ser Trp His Arg Ser Pro Leu Pro Ala Val Val Leu Pro

150

1                      5                      10                      15  
 Ser Val Leu Gln Thr Ala Leu Ser Pro Leu Ala Leu Cys Gln Ala Trp  
                             20                      25                      30  
 5 Arg Arg Ala Val Pro His Gly Val Pro Ser Gln Arg Leu Arg Asn Gln  
                             35                      40                      45  
 10 Glu Ala Ser Leu Val Pro Lys Gly Val Pro Arg Ala Trp Tyr Pro Gly  
                             50                      55                      60  
 Pro Leu Gln Asn Gly Leu Trp Thr His Leu Glu Lys Gly Glu Leu Leu  
                             65                      70                      75                      80  
 15 Gly Leu Lys Pro Thr Pro Gly Gly Leu Leu Leu Leu Arg Ser Phe Trp  
                             85                      90                      95  
 Asp Pro His Pro Ser Arg Pro Phe Leu Cys Thr Leu Leu Pro Pro Pro  
                             100                      105                      110  
 20 Leu Xaa Ile Phe Pro Pro Leu Arg Cys Ser Ala Xaa  
                             115                      120

25

(2) INFORMATION FOR SEQ ID NO: 55:

- (i) SEQUENCE CHARACTERISTICS:  
       (A) LENGTH: 180 amino acids  
 30       (B) TYPE: amino acid  
       (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

35 Met Thr Ser Ala Gly Pro Val Xaa Leu Phe Leu Leu Val Ser Ile Ser  
       1                      5                      10                      15  
 Thr Ser Val Ile Leu Met Gln His Leu Leu Xaa Ala Ser Tyr Cys Asp  
                             20                      25                      30  
 40 Leu Leu His Lys Ala Ala Ala His Leu Gly Cys Trp Gln Lys Val Asp  
                             35                      40                      45  
 Pro Ala Leu Cys Ser Asn Val Leu Gln His Pro Trp Thr Glu Glu Cys  
                             50                      55                      60  
 45 Met Trp Pro Gln Gly Val Leu Val Lys His Ser Lys Asn Val Tyr Lys  
                             65                      70                      75                      80  
 50 Ala Val Gly Xaa Xaa Xaa Val Ala Ile Pro Ser Asp Val Ser His Phe  
                             85                      90                      95  
 Arg Phe Xaa Phe Phe Phe Ser Lys Pro Leu Arg Ile Leu Asn Ile Leu  
                             100                      105                      110  
 55 Leu Leu Leu Glu Gly Ala Val Ile Val Tyr Gln Leu Tyr Ser Leu Met  
                             115                      120                      125  
 Ser Ser Glu Lys Trp His Gln Thr Ile Ser Leu Ala Leu Ile Leu Phe  
                             130                      135                      140  
 60

151

Ser Asn Tyr Tyr Ala Phe Phe Lys Leu Leu Arg Asp Arg Leu Val Leu  
 145 150 155 160

5 Gly Lys Ala Tyr Ser Tyr Ser Ala Ser Pro Gln Arg Asp Leu Asp His  
 165 170 175

Arg Phe Ser Xaa  
 180

10

(2) INFORMATION FOR SEQ ID NO: 56:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 287 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

20 Met Pro Leu Phe Lys Leu Tyr Met Val Met Ser Ala Cys Phe Leu Ala  
 1 5 10 15

Ala Gly Ile Phe Trp Val Ser Ile Leu Cys Arg Asn Thr Tyr Ser Val  
 20 25 30

25 Phe Lys Ile His Trp Leu Met Ala Ala Leu Ala Phe Thr Lys Ser Ile  
 35 40 45

30 Ser Leu Leu Phe His Ser Ile Asn Tyr Tyr Phe Ile Asn Ser Gln Gly  
 50 55 60

Pro Pro His Arg Arg Pro Cys Arg His Val Leu His Arg Thr Pro Ala  
 65 70 75 80

35 Glu Gly Arg Pro Pro Leu His His His Arg Pro Asp Trp Leu Arg Leu  
 85 90 95

Gly Phe Ile Lys Tyr Val Leu Ser Asp Lys Glu Lys Lys Val Phe Gly  
 100 105 110

40 Ile Val Ile Pro Met Gln Val Leu Ala Asn Val Ala Tyr Ile Ile Ile  
 115 120 125

45 Glu Ser Arg Glu Glu Gly Ala Thr Asn Tyr Val Leu Trp Lys Glu Ile  
 130 135 140

Leu Phe Leu Val Asp Leu Ile Cys Cys Gly Ala Ile Leu Phe Pro Val  
 145 150 155 160

50 Val Trp Ser Ile Arg His Leu Gln Asp Ala Ser Gly Thr Asp Gly Lys  
 165 170 175

Val Ala Val Asn Leu Ala Lys Leu Lys Leu Phe Arg His Tyr Tyr Val  
 180 185 190

55 Met Val Ile Cys Tyr Val Tyr Phe Thr Arg Ile Ile Ala Ile Leu Leu  
 195 200 205

60 Gln Val Ala Val Pro Phe Gln Trp Gln Trp Leu Tyr Xaa Leu Leu Val  
 210 215 220

Glu Gly Ser Thr Leu Ala Phe Phe Val Leu Thr Gly Tyr Lys Phe Gln  
 225 230 235 240

5 Pro Thr Gly Asn Asn Pro Tyr Leu Gln Leu Pro Gln Glu Asp Glu Glu  
 245 250 255

Asp Val Gln Met Glu Gln Val Met Thr Asp Ser Gly Phe Arg Glu Gly  
 260 265 270

10

Leu Ser Lys Val Asn Lys Thr Ala Ser Gly Arg Glu Leu Leu Xaa  
 275 280 285

15

(2) INFORMATION FOR SEQ ID NO: 57:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

20

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

25 Met Pro Met Val Phe Leu Leu Leu Phe Asn Leu Met Ser Trp Leu Ile  
 1 5 10 15

Arg Asn Ala Arg Val Ile Leu Arg Ser Leu Asn Leu Lys Arg Asp Gln  
 20 25 30

30 Val Xaa

35

(2) INFORMATION FOR SEQ ID NO: 58:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

40

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

45 Met Lys Ile Val Val Leu Leu Pro Leu Phe Leu Leu Ala Thr Phe Pro  
 1 5 10 15

Arg Lys Leu Gln Thr Cys Leu Xaa  
 20

50

(2) INFORMATION FOR SEQ ID NO: 59:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

55

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

60 Met Ser Gly Gly Glu Gly Ala Ala Leu Pro Ile Leu Leu Leu Leu Leu  
 1 5 10 15

153

Ala Leu Arg Gly Thr Phe His Gly Ala Arg Pro Gly Gly Gly Ala Ser  
20 25 30

5 Gly Ile Trp Cys Leu Leu Leu Pro Glu Gln Glu Pro Pro Val Xaa  
35 40 45

10 (2) INFORMATION FOR SEQ ID NO: 60:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 114 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly  
1 5 10 15

20

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly  
20 25 30

25

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys  
35 40 45

Cys Met Asp Cys Ala Ser Cys Arg Ala Arg Pro His Ser Asp Phe Cys  
50 55 60

30

Leu Gly Cys Ala Ala Ala Pro Pro Ala Pro Phe Arg Leu Leu Trp Pro  
65 70 75 80

Ile Leu Gly Gly Ala Leu Ser Leu Thr Phe Val Leu Gly Leu Leu Ser  
85 90 95

35

Gly Phe Leu Val Trp Arg Arg Cys Arg Arg Glu Arg Ser Ser Pro Pro  
100 105 110

Pro Xaa

40

(2) INFORMATION FOR SEQ ID NO: 61:

45

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

Met Val Cys Ile Leu Val Leu Thr Leu Val Ser Tyr Ser Ser Leu Val  
1 5 10 15

55

Asn Ser Pro Leu Pro Phe Val His Leu Xaa Val Gly Ile Ser Ala Xaa  
20 25 30

60

## (2) INFORMATION FOR SEQ ID NO: 62:

- 5 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 81 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

Met Thr Gly Gly Phe Leu Ser Cys Ile Leu Gly Leu Val Leu Pro Leu  
 1 5 10 15  
 Ala Tyr Xaa Ser Ser Leu Thr Trp Cys Trp Trp Arg Trp Gly Leu Pro  
 15 20 25 30  
 Xaa Pro Ala Gly Pro Pro Arg Cys Thr Pro Gly Cys Asn Ala Ser Gly  
 35 40 45  
 20 Ala Gly Arg Gly Pro Ser Pro Gly Pro Pro Gly Gly Glu Leu His Thr  
 50 55 60  
 Pro Ala Ser Arg Asp Pro Gly Pro Gly Ala Glu Trp Arg Gly Thr Ser  
 65 70 75 80  
 25 Xaa

30

## (2) INFORMATION FOR SEQ ID NO: 63:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 104 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

Met Ala Ala Pro Val Asp Leu Glu Leu Lys Lys Ala Phe Thr Glu Leu  
 40 1 5 10 15  
 Gln Ala Lys Val Ile Asp Thr Gln Gln Lys Val Lys Leu Ala Asp Ile  
 20 25 30  
 45 Gln Ile Glu Gln Leu Asn Arg Thr Lys Lys His Ala His Leu Thr Asp  
 35 40 45  
 Thr Glu Ile Met Thr Leu Val Asp Glu Thr Asn Met Tyr Glu Gly Val  
 50 55 60  
 Gly Arg Met Phe Ile Leu Gln Ser Lys Glu Ala Ile His Ser Gln Leu  
 65 70 75 80  
 55 Leu Glu Lys Gln Lys Ile Ala Glu Glu Lys Ile Lys Glu Leu Glu Gln  
 85 90 95  
 Lys Lys Ser Tyr Leu Glu Arg Arg  
 100

60

## (2) INFORMATION FOR SEQ ID NO: 64:

## 5 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 146 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

10 Met Pro Ser Gly Phe Gln Thr Cys Leu Leu Phe Thr Leu Ser Pro Phe  
 1 5 10 15

Ser Leu Ser Lys Ile Val Gly Val Pro Ser Gln Gln Leu Pro Gly Gln  
 20 25 30

15 Leu Ser Glu Gln Gly Gly Leu Cys Gly His Glu Gly Glu Pro Ala Arg  
 35 40 45

20 Thr Val Pro Glu Thr Gln Leu Pro Leu Pro Phe Asn Ser Ala Gly Pro  
 50 55 60

Pro His Leu Lys Cys Thr Gly Ala Gly Lys Arg Val Trp Ser Pro Pro  
 65 70 75 80

25 Arg Arg Ala Ala Gln Glu Val Ser Leu Gln Leu Val Ser Cys His Pro  
 85 90 95

Cys Arg Gln His Thr Ser Arg Ala Phe Ser Leu Ala Thr Asp Arg Thr  
 100 105 110

30 Ala Ser Ala Arg Val Cys Cys Arg Ser Pro Leu Ser Thr Leu Ile His  
 115 120 125

35 His Thr Arg Gly Gly Gln Arg Cys Arg Glu His Gly Leu Ser Leu Pro  
 130 135 140

Leu Xaa  
 145

40

## (2) INFORMATION FOR SEQ ID NO: 65:

## 45 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

50 Met Ala Ile Leu Met Leu Leu Ala Gly Ser Pro Cys Thr Leu Ser Phe  
 1 5 10 15

Ser Thr Asp Thr Gly Ser Ser Ala Pro Gly Pro Lys Ile Pro Xaa  
 20 25 30

55

## (2) INFORMATION FOR SEQ ID NO: 66:

60

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 260 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

5  
 Met Asp Pro Gln Gly Gln Thr Leu Leu Leu Phe Leu Phe Val Asp Phe  
 1 5 10 15

10  
 His Ser Ala Phe Pro Val Gln Gln Met Glu Ile Trp Gly Val Tyr Thr  
 20 25 30

15  
 Leu Leu Thr Thr His Leu Asn Ala Ile Leu Val Glu Ser His Ser Val  
 35 40 45

20  
 Val Gln Gly Ser Ile Gln Phe Thr Val Asp Lys Val Leu Glu Gln His  
 50 55 60

25  
 His Gln Ala Ala Lys Ala Gln Gln Lys Leu Gln Ala Ser Leu Ser Val  
 65 70 75 80

30  
 Ala Val Asn Ser Ile Met Ser Ile Leu Thr Gly Ser Thr Arg Ser Ser  
 85 90 95

35  
 Phe Arg Lys Met Cys Leu Gln Thr Leu Gln Ala Ala Asp Thr Gln Glu  
 100 105 110

40  
 Phe Arg Thr Lys Leu His Lys Val Phe Arg Glu Ile Thr Gln His Gln  
 115 120 125

45  
 Phe Leu His His Cys Ser Cys Glu Val Lys Gln Leu Thr Leu Glu Lys  
 130 135 140

50  
 Lys Asp Ser Ala Gln Gly Thr Glu Asp Ala Pro Asp Asn Ser Ser Leu  
 145 150 155 160

55  
 Glu Leu Leu Ala Asp Thr Ser Gly Gln Ala Glu Asn Lys Arg Leu Lys  
 165 170 175

60  
 Arg Gly Ser Pro Arg Ile Glu Glu Met Arg Ala Leu Arg Ser Ala Arg  
 180 185 190

65  
 Ala Pro Ser Pro Ser Glu Ala Ala Pro Arg Arg Pro Glu Ala Thr Ala  
 195 200 205

70  
 Ala Pro Leu Thr Pro Arg Gly Arg Glu His Arg Glu Ala His Gly Arg  
 210 215 220

75  
 Ala Leu Ala Pro Gly Arg Ala Ser Leu Gly Ser Arg Leu Glu Asp Val  
 225 230 235 240

80  
 Leu Trp Leu Gln Glu Val Ser Asn Leu Ser Glu Trp Leu Ser Pro Ser  
 245 250 255

85  
 Pro Gly Pro Xaa  
 260

(2) INFORMATION FOR SEQ ID NO: 67:

60

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

Met Ala Ala Ala Cys Gly Pro Gly Ala Ala Gly Thr Ala Cys Ser Ser  
 1 5 10 15

10 Ala Cys Ile Cys Phe Cys Xaa  
 20

## 15 (2) INFORMATION FOR SEQ ID NO: 68:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

Met His Ala Leu Ile Leu Gln Phe Ile Phe Ser Leu Cys Met Tyr Ile  
 1 5 10 15

25 Ser Leu Phe Ser Ala Ala Arg Phe Leu Phe Xaa  
 20 25

## 30 (2) INFORMATION FOR SEQ ID NO: 69:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

40 Leu Leu Leu Leu Cys Phe Cys Cys His Pro Thr His Leu Gln Gly Xaa  
 1 5 10 15

Trp Ala Leu Asp Leu Gly Leu Phe Pro Phe Asn Cys Xaa  
 20 25

## 45 (2) INFORMATION FOR SEQ ID NO: 70:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 216 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

55 Met Tyr Leu Ser Ile Ile Phe Leu Ala Phe Val Ser Ile Asp Arg Cys  
 1 5 10 15

Leu Gln Leu Thr His Ser Cys Lys Ile Tyr Arg Ile Gln Glu Pro Gly  
 20 25 30

60

[illegible]

40 (2) INFORMATION FOR SEQ ID NO: 71:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 407 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

|    |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|    | Met | His | Pro | Ala | Val | Phe | Leu | Ser | Leu | Pro | Asp | Leu | Arg | Cys | Ser | Leu |
|    | 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| 50 | Leu | Leu | Leu | Val | Thr | Trp | Val | Phe | Thr | Pro | Val | Thr | Thr | Glu | Ile | Thr |
|    |     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
|    | Ser | Leu | Asp | Thr | Glu | Asn | Ile | Asp | Glu | Ile | Leu | Asn | Asn | Ala | Asp | Val |
| 55 |     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
|    | Ala | Leu | Val | Asn | Phe | Tyr | Ala | Asp | Trp | Cys | Arg | Phe | Ser | Gln | Met | Leu |
|    |     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| 60 | His | Pro | Ile | Phe | Glu | Glu | Ala | Ser | Asp | Val | Ile | Lys | Glu | Glu | Phe | Pro |
|    | 65  |     |     |     | 70  |     |     |     |     |     | 75  |     |     |     |     | 80  |

[illegible]

Arg Asp Arg Asp Glu Leu Xaa  
405

5

(2) INFORMATION FOR SEQ ID NO: 72:

- 10 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

15 Tyr Leu Ile Ser Tyr Leu Cys Phe Xaa  
1 5

20 (2) INFORMATION FOR SEQ ID NO: 73:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 73:

30 Met Pro Leu Lys Ala Val Thr Trp Pro Thr Leu Asn Ser Lys Leu Val  
1 5 10 15

Ala Ala Val Val Asn Leu Lys Ala Ser Gln Met Pro Ala Ser Ser Arg  
20 25 30

35 Val Xaa

40 (2) INFORMATION FOR SEQ ID NO: 74:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 57 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

Gln Ser Pro Arg Ser Ser Ala Leu Gly Ala Gly Gln Lys Leu Ala Val  
1 5 10 15

50 Cys Ser Pro Asp Ile Leu Cys Cys Pro Thr Asp Thr Leu Leu Ala Ser  
20 25 30

His Pro His Ser Leu Leu Thr Gly Thr Gln Phe Ser Gly Gln Thr Gln  
35 40 45

55 Ala Leu Ala Pro Ser Trp Cys Ala Xaa  
50 55

60

## (2) INFORMATION FOR SEQ ID NO: 75:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

Met Ala Gly Ile His Arg Ala Phe Leu Val Phe Cys Leu Trp Gly Leu  
 1 5 10 15  
 Xaa Leu Cys Val Val Gly Gly Pro Trp Xaa  
 20 25

## (2) INFORMATION FOR SEQ ID NO: 76:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 15 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

Met Ser Phe Ser Ser Pro Lys Ser Leu Leu Ser Leu Ile Ser Xaa  
 1 5 10 15

## (2) INFORMATION FOR SEQ ID NO: 77:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 77:

Met Thr Ile Trp Gln Leu Phe Ala Val Leu Ile Val Leu Phe Ala Lys  
 1 5 10 15  
 Ser Arg Glu Ile Ser Thr Glu Gly Glu Pro Cys Val Leu Ser Lys Asn  
 20 25 30

Xaa

## (2) INFORMATION FOR SEQ ID NO: 78:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

Met Leu Asn Pro Phe Xaa Gln Leu Leu Val Leu Leu Phe Pro Glu  
 1 5 10 15

Trp Pro Thr Pro Leu His Xaa

## 5 (2) INFORMATION FOR SEQ ID NO: 79:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 173 amino acids

(B) TYPE: amino acid

10 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

Met Lys Thr Leu Phe Leu Gly Val Thr Leu Gly Leu Ala Ala Ala Leu  
 1 5 10 15  
 Ser Xaa Thr Leu Xaa Glu Glu Asp Ile Thr Gly Thr Trp Tyr Val Lys  
 20 25 30  
 Ala Met Val Val Asp Lys Thr Phe Arg Arg Gln Glu Ala Gln Lys Val  
 35 40 45  
 Ser Pro Val Lys Val Thr Ala Leu Gly Gly Gly Lys Leu Glu Ala Thr  
 50 55 60  
 Phe Thr Phe Met Arg Glu Asp Arg Cys Ile Gln Lys Lys Ile Leu Xaa  
 65 70 75 80  
 Arg Lys Thr Glu Glu Pro Gly Lys Tyr Ser Ala Cys Glu Pro Leu Pro  
 85 90 95  
 His Ser His Pro His Xaa Pro Pro Pro Thr Pro Val His Gln Pro  
 100 105 110  
 Pro Gln Val Glu Ser Ala Gln Ala Ala Leu Leu Pro Gly Pro Gln Leu  
 115 120 125  
 Cys Pro Pro Pro Arg Arg Gly Trp Pro Leu Leu Pro Gly Gly Leu Val  
 130 135 140  
 Ala Leu Thr Ser Asp Thr Gly Cys Asp Arg Leu Val Arg Ser Arg Asp  
 145 150 155 160  
 Gly Pro Asp His Ala Cys Pro Leu Gly Gly Pro Ser His  
 165 170

45

## (2) INFORMATION FOR SEQ ID NO: 80:

50

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 208 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

Met Ala Asp Ser Ser Tyr Thr Ser Glu Val Gln Ala Ile Leu Ala Phe  
 1 5 10 15  
 Leu Ser Leu Gln Arg Thr Gly Ser Gly Gly Pro Gly Asn His Pro His  
 20 25 30

60

Gly Pro Asp Ala Ser Ala Glu Gly Leu Asn Pro Tyr Gly Leu Val Ala  
 35 40 45  
 5 Pro Arg Phe Gln Arg Lys Phe Lys Ala Lys Gln Leu Thr Pro Arg Ile  
 50 55 60  
 Leu Glu Ala His Gln Asn Val Ala Gln Leu Ser Leu Ala Glu Ala Gln  
 65 70 75 80  
 10 Leu Arg Phe Ile Gln Ala Trp Gln Ser Leu Pro Asp Phe Gly Ile Ser  
 85 90 95  
 Tyr Val Met Val Arg Phe Lys Gly Ser Arg Lys Asp Glu Ile Leu Gly  
 100 105 110  
 15 Ile Ala Asn Asn Arg Leu Ile Arg Ile Asp Leu Ala Val Gly Asp Val  
 115 120 125  
 20 Val Lys Thr Trp Arg Phe Ser Asn Met Arg Gln Trp Asn Val Asn Trp  
 130 135 140  
 Asp Ile Arg Xaa Val Ala Ile Glu Phe Asp Glu His Ile Asn Val Ala  
 145 150 155 160  
 25 Phe Ser Cys Val Ser Ala Ser Cys Arg Ile Val His Glu Tyr Ile Gly  
 165 170 175  
 Gly Tyr Ile Phe Leu Ser Thr Arg Glu Xaa Ala Arg Gly Glu Glu Leu  
 180 185 190  
 30 Asp Glu Asp Leu Phe Leu Gln Leu Thr Gly Gly His Glu Ala Phe Xaa  
 195 200 205

35

40 (2) INFORMATION FOR SEQ ID NO: 81:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

Met Ile Phe Leu Leu Phe Leu Thr Pro Leu Trp Leu Gln Lys Gly Ser  
 1 5 10 15  
 50 Ala Gly Lys Met Ser Gly Glu Phe Leu Tyr Ala Ser Leu Phe Gln Trp  
 20 25 30  
 55 Asn Tyr Phe Trp Arg Asn Lys Lys Val Cys Xaa  
 35 40

60 (2) INFORMATION FOR SEQ ID NO: 82:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 146 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

Met Pro Ser Gly Phe Gln Thr Cys Leu Leu Phe Thr Leu Ser Pro Phe  
 1 5 10 15

10 Ser Leu Ser Lys Ile Val Gly Val Pro Ser Gln Gln Leu Pro Gly Gln  
 20 25 30

Leu Ser Glu Gln Gly Gly Leu Cys Gly His Glu Gly Glu Pro Ala Arg  
 35 40 45

15 Thr Val Pro Glu Thr Gln Leu Pro Leu Pro Phe Asn Ser Ala Gly Pro  
 50 55 60

20 Pro His Leu Lys Cys Thr Gly Ala Gly Lys Arg Val Trp Ser Pro Pro  
 65 70 75 80

Arg Arg Ala Ala Gln Glu Val Ser Leu Gln Leu Val Ser Cys Xaa Pro  
 85 90 95

25 Cys Arg Gln Xaa Thr Ser Arg Ala Phe Ser Leu Ala Thr Asp Arg Thr  
 100 105 110

Ala Ser Ala Arg Val Cys Cys Arg Phe Pro Phe Lys His Thr His Ser  
 115 120 125

30 Pro His Pro Arg Arg Pro Glu Val Gln Gly Ala Trp Ala Val Val Pro  
 130 135 140

Leu Xaa  
 145

35

## 40 (2) INFORMATION FOR SEQ ID NO: 83:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 83:

Met Pro Trp Arg Arg Ala Gly Leu Met Met Leu Pro Ile Ile Thr Gly  
 1 5 10 15

50 Cys Cys Pro Cys Ser Ala Ser Ile Xaa  
 20 25

## 55 (2) INFORMATION FOR SEQ ID NO: 84:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

## 60 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 84:

Met Lys Thr Leu Phe Leu Gly Val Thr Leu Gly Leu Ala Leu Pro Cys  
 1 5 10 15  
 Pro Ser Pro Trp Xaa Arg Arg Ile Ser Gln Gly Pro Gly Thr Xaa  
 20 25 30

10

## (2) INFORMATION FOR SEQ ID NO: 85:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 374 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

Met Ser Val Pro Ala Phe Ile Asp Ile Ser Glu Glu Asp Gln Ala Ala  
 1 5 10 15  
 Glu Leu Arg Ala Tyr Leu Lys Ser Lys Gly Ala Glu Ile Ser Glu Glu  
 20 25 30  
 Asn Ser Glu Gly Gly Leu His Val Asp Leu Ala Gln Ile Ile Glu Ala  
 35 40 45  
 Cys Asp Val Cys Leu Lys Glu Asp Asp Lys Asp Val Glu Ser Val Met  
 50 55 60  
 Asn Ser Val Val Ser Leu Leu Leu Ile Leu Glu Pro Asp Lys Gln Glu  
 65 70 75 80  
 Ala Leu Ile Glu Ser Leu Cys Glu Lys Leu Val Lys Phe Arg Glu Gly  
 85 90 95  
 Glu Arg Pro Ser Leu Arg Leu Gln Leu Leu Ser Asn Leu Phe His Gly  
 100 105 110  
 Met Asp Lys Asn Thr Pro Val Arg Tyr Thr Val Tyr Cys Ser Leu Ile  
 115 120 125  
 Lys Val Ala Ala Ser Cys Gly Ala Ile Gln Tyr Ile Pro Thr Glu Leu  
 130 135 140  
 Asp Gln Val Arg Lys Trp Ile Ser Asp Trp Asn Leu Thr Thr Glu Lys  
 145 150 155 160  
 Lys His Thr Leu Leu Arg Leu Leu Tyr Glu Ala Leu Val Asp Cys Lys  
 165 170 175  
 Lys Ser Asp Ala Ala Ser Lys Val Met Val Glu Leu Leu Gly Ser Tyr  
 180 185 190  
 Thr Glu Asp Asn Ala Ser Gln Ala Arg Val Asp Ala His Arg Cys Ile  
 195 200 205  
 Val Arg Ala Leu Lys Asp Pro Asn Ala Phe Leu Phe Asp His Leu Leu  
 210 215 220

60

Thr Leu Lys Pro Val Lys Phe Leu Glu Gly Glu Leu Ile His Asp Leu  
 225 230 235 240  
 5 Leu Thr Ile Phe Val Ser Ala Lys Leu Ala Ser Tyr Val Lys Phe Tyr  
 245 250 255  
 Gln Asn Asn Lys Asp Phe Ile Asp Ser Leu Gly Leu Leu His Glu Gln  
 260 265 270  
 10 Asn Met Ala Lys Met Arg Leu Leu Thr Phe Met Gly Met Ala Val Glu  
 275 280 285  
 Asn Lys Glu Ile Ser Phe Asp Thr Met Gln Gln Glu Leu Gln Ile Gly  
 290 295 300  
 15 Ala Asp Asp Val Glu Ala Phe Val Ile Asp Ala Val Arg Thr Lys Met  
 305 310 315 320  
 Val Tyr Cys Lys Ile Asp Gln Thr Gln Arg Lys Val Val Val Ser His  
 325 330 335  
 Ser Thr His Arg Thr Phe Gly Lys Gln Gln Trp Gln Gln Leu Tyr Asp  
 340 345 350  
 25 Thr Leu Asn Ala Trp Lys Gln Asn Leu Asn Lys Val Lys Asn Ser Leu  
 355 360 365  
 Leu Ser Leu Ser Asp Thr  
 370  
 30

(2) INFORMATION FOR SEQ ID NO: 86:

35 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 86:  
 Met Ser Val Pro Ala Phe Ile Asp Ile Ser Glu Glu Asp  
 1 5 10

45

(2) INFORMATION FOR SEQ ID NO: 87:

50 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 15 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

55 Gln Ala Ala Glu Leu Arg Ala Tyr Leu Lys Ser Lys Gly Ala Glu  
 1 5 10 15

60

(2) INFORMATION FOR SEQ ID NO: 88:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

Ile Ser Glu Glu Asn Ser Glu Gly Gly Leu His Val Asp Leu Ala Gln  
1 5 10 15

10 Ile

## 15 (2) INFORMATION FOR SEQ ID NO: 89:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

Ile Glu Ala Cys Asp Val Cys Leu Lys Glu Asp Asp Lys Asp Val Glu  
1 5 10 15

25 Ser Val

30

## (2) INFORMATION FOR SEQ ID NO: 90:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

Val Ala Arg Pro Ser Ser Leu Phe Arg Ser Ala Trp Ser Cys Glu Trp  
1 5 10 15

40

45

## (2) INFORMATION FOR SEQ ID NO: 91:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 91:

Leu Arg Leu Gln Leu Leu Ser Asn Leu Phe His Gly  
1 5 10

55

## 60 (2) INFORMATION FOR SEQ ID NO: 92:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

Lys Asp Val Glu Ser Val Met Asn Ser Val Val Ser Leu Leu Leu Ile  
 1 5 10 15  
 Leu

## (2) INFORMATION FOR SEQ ID NO: 93:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

Asp Ala Ala Ser Lys Val Met Val Glu Leu Leu Gly Ser Tyr Thr Glu  
 1 5 10 15  
 Asp Asn Ala Ser Gln Ala Arg Val Asp Ala  
 20 25

## (2) INFORMATION FOR SEQ ID NO: 94:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 10 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

Val Glu Ala Phe Val Ile Asp Ala Val Arg  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO: 95:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 35 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

Met Ser Glu Ile Tyr Leu Arg Cys Gln Asp Glu Gln Gln Tyr Ala Arg  
 1 5 10 15  
 Trp Met Ala Gly Cys Arg Leu Ala Ser Lys Gly Arg Thr Met Ala Asp  
 20 25 30  
 Ser Ser Tyr  
 35

## (2) INFORMATION FOR SEQ ID NO: 96:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

Leu Val Ala Pro Arg Phe Gln Arg Lys Phe Lys Ala Lys Gln Leu Thr  
 1 5 10 15  
 Pro Arg Ile Leu Glu Ala His Gln Asn Val Ala Gln Leu Ser Leu Ala  
 20 25 30  
 Glu Ala Gln Leu Arg Phe Ile Gln Ala Trp Gln Ser Leu  
 35 40 45

## (2) INFORMATION FOR SEQ ID NO: 97:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

Val Gly Asp Val Val Lys Thr Trp Arg Phe Ser Asn Met Arg Gln Trp  
 1 5 10 15  
 Asn Val Asn Trp Asp Ile Arg  
 20

## (2) INFORMATION FOR SEQ ID NO: 98:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

Glu Glu Ile Asp Cys Thr Glu Glu Glu Met Met Val Phe Ala Ala Leu  
 1 5 10 15  
 Gln Tyr His Ile Asn Lys Leu Ser Gln Ser  
 20 25

## (2) INFORMATION FOR SEQ ID NO: 99:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

Glu Glu Ile Asp Cys Thr Glu Glu Glu Met Met Val Phe Ala Ala Leu  
1 5 10 15

Gln Tyr His Ile Asn Lys Leu Ser Gln Ser  
20 25

10

(2) INFORMATION FOR SEQ ID NO: 100:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

Lys Glu Leu Ser Phe Ala Arg Ile Lys Ala Val Glu Cys Val Glu Ser  
1 5 10 15

Thr Gly Arg His Ile Tyr Phe Thr Leu Val  
20 25

25

(2) INFORMATION FOR SEQ ID NO: 101:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

30

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

Gly Trp Asn Ala Gln Ile Thr Leu Gly Leu Val Lys Phe Lys Asn Gln  
1 5 10 15

Gln

40

(2) INFORMATION FOR SEQ ID NO: 102:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 16 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

Leu Val Leu Gly Leu Ser Xaa Leu Asn Asn Ser Tyr Asn Phe Ser Phe  
1 5 10 15

55

(2) INFORMATION FOR SEQ ID NO: 103:

60

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

His Val Val Ile Gly Ser Gln Ala Glu Glu Gly Gln Tyr Ser Leu Asn  
 1 5 10 15

10 Phe

## 15 (2) INFORMATION FOR SEQ ID NO: 104:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

His Asn Cys Asn Asn Ser Val Pro Gly Lys Glu His Pro Phe Asp Ile  
 1 5 10 15

25 Thr Val Met

30

## (2) INFORMATION FOR SEQ ID NO: 105:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 17 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 35 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

Phe Ile Lys Tyr Val Leu Ser Asp Lys Glu Lys Lys Val Phe Gly Ile  
 1 5 10 15

40 Val

45

## (2) INFORMATION FOR SEQ ID NO: 106:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 13 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## 50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

Ile Pro Met Gln Val Leu Ala Asn Val Ala Tyr Ile Ile  
 1 5 10

## 60 (2) INFORMATION FOR SEQ ID NO: 107:

- 5 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:
- 10 Ile Pro Met Gln Val Leu Ala Asn Val Ala Tyr Ile Ile  
1 5 10
- 15 (2) INFORMATION FOR SEQ ID NO: 108:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 108:
- 20 Asp Gly Lys Val Ala Val Asn Leu Ala Lys Leu Lys Leu Phe Arg  
1 5 10 15
- 25 (2) INFORMATION FOR SEQ ID NO: 109:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 109:
- 35 Ile Arg Glu Lys Asn Pro Asp Gly Phe Leu Ser Ala Ala  
1 5 10
- 40 (2) INFORMATION FOR SEQ ID NO: 110:
- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 110:
- 45 Met Met Phe Gly Gly Tyr Glu Thr Ile  
1 5
- 50 (2) INFORMATION FOR SEQ ID NO: 111:
- 55 (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 111:
- 60 Tyr Arg Asp Glu Ser Ser Ser Glu Leu Ser Val Asp Ser Glu Val Glu

[illegible]

5

(2) INFORMATION FOR SEQ ID NO: 112:

10

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 136 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

Tyr Ala Gln Asp Leu Asp Asp Val Ile Arg Glu Glu Glu His Glu Glu  
1 5 10 15

20 Lys Asn Ser Gly Asn Ser Glu Ser Ser Ser Ser Lys Pro Asn Gln Lys  
20 25 30

Lys Leu Ile Val Leu Ser Asp Ser Glu Val Ile Gln Leu Ser Asp Gly  
35 40 45

25 Ser Glu Val Ile Thr Leu Ser Asp Glu Asp Ser Ile Tyr Arg Cys Lys  
50 55 60

Gly Lys Asn Val Arg Val Gln Ala Gln Glu Asn Ala His Gly Leu Ser  
65 70 75 80

Ser Ser Leu Gln Ser Asn Glu Leu Val Asp Lys Lys Cys Lys Ser Asp  
85 90 95

35 Ile Glu Lys Pro Lys Ser Glu Glu Arg Ser Gly Val Ile Arg Glu Val  
100 105 110

Met Ile Ile Glu Val Ser Ser Ser Glu Glu Glu Glu Ser Thr Ile Ser  
115 120 125

40      Glu Gly Asp Asn Val Glu Ser Trp  
              130                          135

45 (2) INFORMATION FOR SEQ ID NO: 113:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 37 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

Met Leu Leu Gly Cys Glu Val Asp Asp Lys Asp Asp Asp Ile Leu Leu  
1 5 10 15

Asn Leu Val Gly Cys Glu Asn Ser Val Thr Glu Gly Glu Asp Gly Ile  
20 25 30

60 Asn Trp Ser Ile Ser  
35

## (2) INFORMATION FOR SEQ ID NO: 114:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

Asp Lys Asp Ile Glu Ala Gln Ile Ala Asn Asn Arg Thr Pro Gly Arg  
 1 5 10 15

Trp Thr

## (2) INFORMATION FOR SEQ ID NO: 115:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

Gln Arg Tyr Tyr Ser Ala Asn Lys Asn Ile Ile Cys Arg Asn Cys Asp  
 1 5 10 15

Lys Arg Gly His Leu Ser Lys Asn Cys Pro Leu Pro Arg Lys Val  
 20 25 30

## (2) INFORMATION FOR SEQ ID NO: 116:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 179 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

Arg Arg Cys Phe Leu Cys Ser Arg Arg Gly His Leu Leu Tyr Ser Cys  
 1 5 10 15

Pro Ala Pro Leu Cys Glu Tyr Cys Pro Val Pro Lys Met Leu Asp His  
 20 25 30

Ser Cys Leu Phe Arg His Ser Trp Asp Lys Gln Cys Asp Arg Cys His  
 35 40 45

Met Leu Gly His Tyr Thr Asp Ala Cys Thr Glu Ile Trp Arg Gln Tyr  
 50 55 60

His Leu Thr Thr Lys Pro Gly Pro Pro Lys Lys Pro Lys Thr Pro Ser  
 65 70 75 80

Arg Pro Ser Ala Leu Ala Tyr Cys Tyr His Cys Ala Gln Lys Gly His  
 85 90 95

Tyr Gly His Glu Cys Pro Glu Arg Glu Val Tyr Asp Pro Ser Pro Val  
 100 105 110  
 5 Ser Pro Phe Ile Cys Tyr Tyr Xaa Asp Lys Tyr Glu Ile Gln Glu Arg  
 115 120 125  
 Glu Lys Arg Leu Lys Gln Lys Ile Lys Val Xaa Lys Lys Asn Gly Val  
 130 135 140  
 10 Ile Pro Glu Pro Ser Lys Leu Pro Tyr Ile Lys Ala Ala Asn Glu Asn  
 145 150 155 160  
 15 Pro His His Asp Ile Arg Lys Gly Arg Ala Ser Trp Lys Ser Asn Arg  
 165 170 175  
 Trp Pro Gln

20

(2) INFORMATION FOR SEQ ID NO: 117:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 17 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

30 Leu Ser Ile Ile Phe Leu Ala Phe Val Ser Ile Asp Arg Cys Leu Gln  
 1 5 10 15  
 Leu

35

(2) INFORMATION FOR SEQ ID NO: 118:

- 40 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 67 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

45 Gly Ser Cys Phe Ala Thr Trp Ala Phe Ile Gln Lys Asn Thr Asn His  
 1 5 10 15  
 50 Arg Cys Val Ser Ile Tyr Leu Ile Asn Leu Leu Thr Ala Asp Phe Leu  
 20 25 30  
 Leu Thr Leu Ala Leu Pro Val Lys Ile Val Val Asp Leu Gly Val Ala  
 35 40 45  
 55 Pro Trp Lys Leu Lys Ile Phe His Cys Gln Val Thr Ala Cys Leu Ile  
 50 55 60

Tyr Ile Asn  
 65

60

## (2) INFORMATION FOR SEQ ID NO: 119:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

Ala Pro Leu Glu Thr Met Gln Asn Lys Pro Arg Ala Pro Gln Lys Arg  
 1 5 10 15  
 Ala Leu Pro Phe Pro Glu Leu Glu Leu Arg Asp Tyr Ala Ser Val Leu  
 20 25 30  
 Thr Arg Tyr Ser Leu Gly Leu Arg Asn Lys Glu Pro Ser Leu Gly His  
 35 40 45  
 Arg Trp Gly Thr Gln Lys Leu Gly Arg Ser Pro Cys  
 50 55 60

## (2) INFORMATION FOR SEQ ID NO: 120:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 166 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

Asn Arg Glu Arg Gly Gly Ala Gly Ala Thr Phe Glu Cys Asn Ile Cys  
 1 5 10 15  
 Leu Glu Thr Ala Arg Glu Ala Val Val Ser Val Cys Gly His Leu Tyr  
 20 25 30  
 Cys Trp Pro Cys Leu His Gln Trp Leu Glu Thr Arg Pro Glu Arg Gln  
 35 40 45  
 Glu Cys Pro Val Cys Lys Ala Gly Ile Ser Arg Glu Lys Val Val Pro  
 50 55 60  
 Leu Tyr Gly Arg Gly Ser Gln Lys Pro Gln Asp Pro Arg Leu Lys Thr  
 65 70 75 80  
 Pro Pro Arg Pro Gln Gly Gln Arg Pro Ala Pro Glu Ser Arg Gly Gly  
 85 90 95  
 Phe Gln Pro Phe Gly Asp Thr Gly Gly Phe His Phe Ser Phe Gly Val  
 100 105 110  
 Gly Ala Phe Pro Phe Gly Phe Phe Thr Thr Val Phe Asn Ala His Glu  
 115 120 125  
 Pro Phe Arg Arg Gly Thr Gly Val Asp Leu Gly Gln Gly His Pro Ala  
 130 135 140  
 Ser Ser Trp Gln Asp Ser Leu Phe Leu Phe Leu Ala Ile Phe Phe Phe  
 145 150 155 160  
 Phe Trp Leu Leu Ser Ile  
 165

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

|  |                            |
|--|----------------------------|
| <b>A.</b> The indications made below relate to the microorganism referred to in the description<br>on page <u>29</u> , line <u>N/A</u>   |                            |
| <b>B. IDENTIFICATION OF DEPOSIT</b> <span style="float: right;">Further deposits are identified on an additional sheet <input type="checkbox"/></span>   |                            |
| Name of depositary institution<br><div style="text-align: center;">American Type Culture Collection</div>  |                            |
| Address of depositary institution (including postal code and country)<br>10801 University Boulevard<br>Manassas, Virginia 20110-2209<br>United States of America   |                            |
| Date of deposit    May 22, 1997  | Accession Number    209075 |
| <b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable) <span style="float: right;">This information is continued on an additional sheet <input type="checkbox"/></span>  |                            |
|  |                            |
| <b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if the indications are not for all designated States)  |                            |
| EUROPE<br><br>In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4)EPC). |                            |
| <b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)   |                            |
| The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")  |                            |
|  |                            |

|  |   |
|--|---|
| <div style="text-align: center; border-bottom: 1px solid black; margin-bottom: 5px;">For receiving Office use only</div> <div style="display: flex; align-items: center; margin-bottom: 10px;"><input checked="checked" style="margin-right: 10px;" type="checkbox"/><span>This sheet was received with the international application</span></div> <div style="border-top: 1px solid black; padding-top: 10px;">Authorized officer<br/><div style="text-align: center; margin-top: 10px;"><b>JERYL McDOWELL</b><br/>703-305-3639</div></div> | <div style="text-align: center; border-bottom: 1px solid black; margin-bottom: 5px;">For International Bureau use only</div> <div style="display: flex; align-items: center; margin-bottom: 10px;"><input style="margin-right: 10px;" type="checkbox"/><span>This sheet was received by the International Bureau on:</span></div> <div style="border-top: 1px solid black; padding-top: 10px;">Authorized officer</div> |
|--|---|

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

|  |  |
|--|--|
| A. The indications made below relate to the microorganism referred to in the description<br>on page <u>30</u> , line <u>N/A</u>  |  |
| <b>B. IDENTIFICATION OF DEPOSIT</b> Further deposits are identified on an additional sheet <input type="checkbox"/>  |  |
| Name of depositary institution<br>American Type Culture Collection   |  |
| Address of depositary institution (including postal code and country)<br>10801 University Boulevard<br>Manassas, Virginia 20110-2209<br>United States of America   |  |
| Date of deposit May 8, 1997  | Accession Number 209022  |
| <b>C. ADDITIONAL INDICATIONS</b> (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>   |  |
| <br><br><br><br><br>   |  |
| <b>D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE</b> (if the indications are not for all designated States)  |  |
| EUROPE<br><br>In respect of those designations in which a European Patent is sought a sample of the deposited microorganism will be available until the publication of the mention of the grant of the European patent or until the date on which the application has been refused or withdrawn or is deemed to be withdrawn, only by the issue of such a sample to an expert nominated by the person requesting the sample (Rule 28(4)EPC). |  |
| <b>E. SEPARATE FURNISHING OF INDICATIONS</b> (leave blank if not applicable)   |  |
| The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")<br><br><br><br><br><br><br><br><br><br>  |  |
| <b>For receiving Office use only</b><br><br><input checked="" type="checkbox"/> This sheet was received with the international application<br><br>Authorized officer<br><b>JERYL McDOWELL</b><br><b>703-305-3636</b>   | <b>For International Bureau use only</b><br><br><input type="checkbox"/> This sheet was received by the International Bureau on:<br><br>Authorized officer |

*What Is Claimed Is:*

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
  - (f) a polynucleotide which is a variant of SEQ ID NO:X;
  - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
  - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
  - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

10

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

15

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

20

9. A recombinant host cell produced by the method of claim 8.

10. The recombinant host cell of claim 9 comprising vector sequences.

25

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

30

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

35

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

- (g) a variant of SEQ ID NO:Y;
- (h) an allelic variant of SEQ ID NO:Y; or
- (i) a species homologue of the SEQ ID NO:Y.

5        12.     The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

10        13.     An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

14.     A recombinant host cell that expresses the isolated polypeptide of claim 11.

15        15.     A method of making an isolated polypeptide comprising:  
      (a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and  
      (b) recovering said polypeptide.

20        16.     The polypeptide produced by claim 15.

17.     A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

25        18.     A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

- (a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and
  - (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.
- 30

19.     A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

- (a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and
  - (b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.
- 35

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:

- 5 (a) contacting the polypeptide of claim 11 with a binding partner; and  
(b) determining whether the binding partner effects an activity of the polypeptide.

21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.


10 22. A method of identifying an activity in a biological assay, wherein the method comprises:

- (a) expressing SEQ ID NO:X in a cell;  
(b) isolating the supernatant;  
(c) detecting an activity in a biological assay; and  
15 (d) identifying the protein in the supernatant having the activity.

23. The product produced by the method of claim 22.

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/10868

| <b>A. CLASSIFICATION OF SUBJECT MATTER</b>   |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
|--|--|--|--|---|--|--|--|--|---|---|--|--|--|--|
| IPC(6) :C07K 1/00; C07H 21/04<br>US CL :530/350; 536/23.5<br>According to International Patent Classification (IPC) or to both national classification and IPC   |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| <b>B. FIELDS SEARCHED</b>  |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| Minimum documentation searched (classification system followed by classification symbols)<br>U.S. : 530/350; 536/23.5  |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)<br>MPSRCH   |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| <b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b>  |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| Category*  | Citation of document, with indication, where appropriate, of the relevant passages   | Relevant to claim No.  |  |   |  |  |  |  |   |   |  |  |  |  |
| X<br>--<br>Y   | Database Genbank on MPSRCH, University of Edinburgh, (Edinburgh, UK), No. N20562, HILLIER et al. 'yx39a08.s1 Homo sapiens cDNA clone 264086 3'.' 18 December 1995, compare to SEQ ID No. 11.   | 1<br>-----<br>2-10, 14, 15, 21   |  |   |  |  |  |  |   |   |  |  |  |  |
| X<br>--<br>Y   | WO 95/31544 A1 (H WEINWURZEL, H.) 23 November 1995, compare Figure 1b to SEQ ID No. 12.  | 1<br>--<br>2-10, 14, 15, 21  |  |   |  |  |  |  |   |   |  |  |  |  |
| X<br>--<br>Y   | Database Genbank on MPSRCH, University of Edinburgh, (Edinburgh, UK), No. N23080, HILLIER et al. 'yw43d02.s1 Homo sapiens cDNA clone 254979 3'.' 28 December 1995, compare to SEQ ID No. 13.   | 1<br>-----<br>2-10, 14, 15, 21   |  |   |  |  |  |  |   |   |  |  |  |  |
| <input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input type="checkbox"/> See patent family annex.   |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| <table border="0"> <tr> <td>* Special categories of cited documents:</td> <td>*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention</td> </tr> <tr> <td>*A* document defining the general state of the art which is not considered to be of particular relevance</td> <td>*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone</td> </tr> <tr> <td>*B* earlier document published on or after the international filing date</td> <td>*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art</td> </tr> <tr> <td>*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)</td> <td>*A* document member of the same patent family</td> </tr> <tr> <td>*O* document referring to an oral disclosure, use, exhibition or other means</td> <td></td> </tr> <tr> <td>*P* document published prior to the international filing date but later than the priority date claimed</td> <td></td> </tr> </table> |  |  | * Special categories of cited documents: | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention | *A* document defining the general state of the art which is not considered to be of particular relevance | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone | *B* earlier document published on or after the international filing date | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art | *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) | *A* document member of the same patent family | *O* document referring to an oral disclosure, use, exhibition or other means |  | *P* document published prior to the international filing date but later than the priority date claimed |  |
| * Special categories of cited documents:   | *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| *A* document defining the general state of the art which is not considered to be of particular relevance   | *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone   |  |  |   |  |  |  |  |   |   |  |  |  |  |
| *B* earlier document published on or after the international filing date   | *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art |  |  |   |  |  |  |  |   |   |  |  |  |  |
| *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  | *A* document member of the same patent family  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| *O* document referring to an oral disclosure, use, exhibition or other means   |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| *P* document published prior to the international filing date but later than the priority date claimed   |  |  |  |   |  |  |  |  |   |   |  |  |  |  |
| Date of the actual completion of the international search<br>02 OCTOBER 1998   |  | Date of mailing of the international search report<br>28 OCT 1998  |  |   |  |  |  |  |   |   |  |  |  |  |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. (703) 305-3230  |  | Authorized officer<br><br>BRUCE CAMPPELL<br>Telephone No. (703) 308-0196 |  |   |  |  |  |  |   |   |  |  |  |  |

## INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US98/10868

| C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT |   |                                |
|---|---|--------------------------------|
| Category*   | Citation of document, with indication, where appropriate, of the relevant passages  | Relevant to claim No.          |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. G23170, HUDSON, T. 'human STS WI-<br>16915', 31 May 1996, compare with SEQ ID No. 14.  | 1<br>-----<br>2-10, 14, 15, 21 |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. H18098, HILLIER et al. 'yn47d01.s1 Homo<br>sapiens cDNA clone 171553 3'' 29 June 1995, compare with SEQ<br>ID No. 15.  | 1<br>-----<br>2-10, 14, 15, 21 |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. N46256, HILLIER et al. 'yy72g09.s1 Homo<br>sapiens cDNA clone 279136 3'' 14 February 1996, compare with<br>SEQ ID No. 16.  | 1<br>-----<br>2-10, 14, 15, 21 |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. N28611, HILLIER et al. 'yx38f03.r1 Homo<br>sapiens cDNA clone 264029 5'' 04 January 1996, compare with<br>SEQ ID No. 17.   | 1<br>-----<br>2-10, 14, 15, 21 |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. R70283, HILLIER et al. 'yj81c08.r1 Homo<br>sapiens cDNA clone 155150 5'' 01 June 1995, compare with SEQ<br>ID No. 18.  | 1<br>---<br>2-10, 14, 15, 21   |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. T98012, HILLIER et al. 'ye56e07.s1 Homo<br>sapiens cDNA clone 121764 3'' 29 March 1995, compare with<br>SEQ ID No. 19.   | 1<br>-----<br>2-10, 14, 15, 21 |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. Z44692, GENEXPRESS. 'H. sapiens partial<br>cDNA sequence; clone 27b07, mRNA sequence.' 21 September<br>1995, compare with SEQ ID No. 20.                                     | 1<br>-----<br>2-10, 14, 15, 21 |
| X<br>--<br>Y  | Database Genbank on MPSRCH, University of Edinburgh,<br>(Edinburgh, UK), No. W83277, MARRA et al. 'mf25e5.r1 Soares<br>mouse embryo NbME13.5 14.5 Mus musculus cDNA clone 406112<br>5', mRNA sequence.' 12 September 1996, compare with SEQ ID<br>No. 43. | 1<br>-----<br>2-10, 14, 15, 21 |

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/10868

## BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

### Group I:

Claims 1-10, 14, 15, and 21 drawn to a polynucleotide(s), vector(s) containing the polynucleotide, host cells containing the vector(s) which are SEQ ID NO: X or a polynucleotide encoding the polypeptide Y or a cDNA in the material deposited with American Type Culture Collection with accession number Z wherein the cDNA in Z hybridizes to X. Additionally Group I contains the first method making the cells (claim 14) containing the vector(s) containing the polynucleotide(s) and the first method of use of the cells (claim 15) to make a product. There appear to be a total of 46 polynucleotide sequences of which the first ten (10) are selected for examination and therefore, there are nine (9) remaining additional groups of four (4) polynucleotide sequences.

### Group II:

Claims 11, 12, 16, and 23 drawn to polypeptides and/or fragments thereof with the amino acid sequence defined by SEQ ID NO: Y as found in the material deposited with the American Type Culture Collection with accession number Z. There appear to be a total of 74 polypeptide sequences and therefore 73 additional species of proteins.

### Group III:

Claim 13, drawn to an antibody and/or fragments thereof that bind to a polypeptide with the amino acid sequence defined by SEQ ID NO: Y as found in the material deposited with the American Type Culture Collection with accession number Z. There appear to be a total of 74 antibodies that correspond to the SEQ ID NOs: for the "Y" and "Z" sequences and therefore 73 additional species of proteins.

### Group IV:

Claim 17, drawn to a process of preventing, treating, or ameliorating a medical condition by administering a polypeptide or a polynucleotide which a second/alternative process of use of the second product and of an alternative process of use of the first claimed product in Group I.

In Group IV, and where additional fees are paid, the claims are searched only insofar as they are applicable to the selected polypeptide and its corresponding SEQ ID NO: as the first species as directed to a process practiced using a polypeptide. The second species is the practice of the process using a polynucleotide. In each instance, the same selected polypeptide as for the first species of Group II and for the first 10 polynucleotide sequences for Group I would be examined. Applicant may elect to pay additional fees for each additional one of the 73 different polypeptide species beyond the first one (1) polypeptide and/or the first 10 polynucleotides as set forth in the above paragraphs directed to Group I and II.

### Group V:

Claim 18, drawn to a method of diagnosis of a pathological condition and another alternative process of use of the first claimed product in Group I. Additionally Group V contains indicia that there are a total of 46 polynucleotide sequences and therefore, nine(9) additional groups of four (4) polynucleotide sequences beyond the first ten (10) sequences.

### Group VI:

Claim 19, drawn to a method of diagnosis of a pathological condition and another alternative process of use of the polypeptide. There appear to be a total of 74 polypeptide sequences and therefore 73 additional species of proteins.

### Group VII:

Claim 20, drawn to a method of identification of a binding partner for a polypeptide. There appear to be a total of 74 polypeptide sequences and therefore 73 additional species of proteins.

### Group VIII:

Claim 22, drawn to a method of identification of function of a protein is another alternative process of use of the product in Group I. Additionally Group V contains indicia that there are a total of 46 polynucleotide sequences and therefore, nine(9) additional groups of four (4) polynucleotide sequences beyond the first ten (10) sequences.

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US98/10868

The inventions listed as Groups I through VIII do not relate to a single inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons.

Claims of Group I are drawn to nucleotides, nucleotide constructs, and/or methods requiring the use of nucleotides or nucleotide constructs that contain more than ten individual, independent, and distinct nucleotide sequences in alternative form. Accordingly, these claims are subject to lack of unity as outlined in 1192 O.G. 68 (19 November 1996).

For Group I, the first ten (10) of the individual polynucleotide sequences designated as "X" by SEQ ID NO: as set forth in the application (see for example page 29+ and/or the SEQUENCE LISTING) are included for search. The corresponding SEQ ID NO: for "Y" and "Z" for each selected "X" should also be noted. The search of the no more than ten sequences may include the complements of the selected sequences and, where appropriate, may include subsequences within the selected sequences (e.g., oligomeric probes and/or primers).

In Group IV (as directed to the species which are polynucleotides) should applicant pay the additional fee for the second appearing species in Group IV which are polynucleotides, first ten (10) of the individual polynucleotide sequences designated as "X" by SEQ ID NO: as set forth in the application (see for example page 29+ and/or the SEQUENCE LISTING) are included for search of Group IV should the fees for Group IV be paid. This is also applied to Groups V and VIII. The corresponding SEQ ID NO: for "Y" and "Z" for each selected "X" should also be noted. The search of the no more than ten sequences may include the complements of the selected sequences and, where appropriate, may include subsequences within the selected sequences (e.g., oligomeric probes and/or primers).

Where Applicant may elect to pay additional fees for a search of sequences beyond the initial ten (10) polynucleotide sequences, and in accordance with 1192 O.G. 68 (19 November 1996), applicant may select additional groups of polynucleotides consisting of four (4) sequences beyond the initial ten (10) sequences for Group I which would then be searched with Group I upon payment of the requisite fees for the requisite Groups beyond Group I.

As to the polypeptides of Groups II, III, IV (as directed to a species which is a polypeptide), VI, and VII each is a distinct and different protein. Should additional fees for the above indicated Groups be paid, the first amino acid sequence identified from the SEQUENCE LISTING by applicant would be searched with the additional group for which the additional search fees were paid.

Applicant may select additional proteins and or antibodies to be searched by specifying the appropriate SEQ ID NOs and payment of the requisite additional fees for each single additional particular species that are selected beyond the one (1) protein identified by SEQ ID NO:.

The SEQ ID NOs in Group I define, absent evidence to the contrary, structurally distinct and different proteins. Note the present application written description (page 5+) refers to the protein encoded by gene 1 as likely to be involved in promotion of a variety of cancers whereas gene 2 (pages 6-7) is directed to apparently a variety but not correlated immune system disorder(s) whereas gene 3 (pages 7-8) is asserted at page 7 to be a mediator of ligand dependent AF-2. Each of which and absent factual evidence to the contrary, are directed to genes encoding distinct and different proteins and are therefore distinct and different genes and appear to map to different chromosomes.

As to the protein of Group II and the antibody of Group III, each is distinct and different for the reasons indicated in the preceding paragraph and because the proteins have distinct and different chemical, physical, and biological properties from that of DNA/polynucleotides/vectors and cells containing same.

Groups IV through VIII are directed to alternative processes of use of the Group I and II compositions where Group I contains in claims 14 and 15, the first claimed method of making the polynucleotide and the first claimed process of use of the cells containing the vector which contains the polynucleotides.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number  
WO 01/45730 A2

(51) International Patent Classification<sup>7</sup>: A61K 38/17,  
39/395, C07K 14/705, 16/28, C12N 15/12, A61P 35/00

(74) Agent: SMITH, Julie, K.; 51 University Street, Seattle,  
WA 98101 (US).

(21) International Application Number: PCT/US00/34755

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(22) International Filing Date:  
19 December 2000 (19.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
60/172,878 20 December 1999 (20.12.1999) US  
60/203,347 10 May 2000 (10.05.2000) US

(84) Designated States (*regional*): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

(71) Applicant (*for all designated States except US*): IM-  
MUNEX CORPORATION [US/US]; 51 University  
Street, Seattle, WA 98101 (US).

**Published:**

— Without international search report and to be republished  
upon receipt of that report.

(72) Inventor; and

(75) Inventor/Applicant (*for US only*): WILEY, Steven, R.  
[US/US]; 1511 - 11th Avenue West #6, Seattle, WA 98119  
(US).

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*



WO 01/45730 A2

(54) Title: TWEAK RECEPTOR

(57) Abstract: The present invention provides the TWEAK receptor and methods for identifying and using agonists and antagonists of the TWEAK receptor. In particular, the invention provides methods of screening for agonists and antagonists and for treating diseases or conditions mediated by angiogenesis, such as solid tumors and vascular deficiencies of cardiac or peripheral tissue.

**TITLE**  
**TWEAK RECEPTOR**

**REFERENCE TO RELATED APPLICATIONS**

5        This application claims the benefit of U.S. Provisional Application Serial Number 60/172,878, filed 20 December 1999, and U.S. Provisional Application Serial Number 60/203,347, filed 10 May 2000, both of which are incorporated herein by reference.

**FIELD OF THE INVENTION**

10        The present invention relates to the discovery of the functional receptor (TWEAKR) for the TWEAK protein. More particularly, the invention relates to the use of TWEAKR agonists and antagonists in methods of treatment, and to screening methods based on TWEAKR and the TWEAK-TWEAKR interaction.

**BACKGROUND OF THE INVENTION**

15        **A. Angiogenesis**

          Angiogenesis is a multi-step developmental process that results in the formation of new blood vessels off of existing vessels. This spatially and temporally regulated process involves loosening of matrix contacts and support cell interactions in the existing vessels by proteases, followed by coordinated  
20        movement, morphological alteration, and proliferation of the smooth muscle and endothelial cells of the existing vessel. The nascent cells then extend into the target tissue followed by cell-cell interactions in which the endothelial cells form tubes which the smooth muscle cells surround. In a coordinated fashion, extracellular matrix proteins of the vessel are secreted and peri-endothelial support cells are recruited to support and maintain structural integrity (see, e.g., Daniel et al., Ann. Rev. Physiol. 2000(62):649, 2000).  
25        Angiogenesis plays important roles in both normal and pathological physiology.

          Under normal physiological conditions, angiogenesis is involved in fetal and embryonic development, wound healing, organ regeneration, and female reproductive remodeling processes including formation of the endometrium, corpus luteum, and placenta. Angiogenesis is stringently regulated under normal conditions, especially in adult animals, and perturbation of the regulatory controls can lead to  
30        pathological angiogenesis.

          Pathological angiogenesis has been implicated in the manifestation and/or progression of inflammatory diseases, certain eye disorders, and cancer. In particular, several lines of evidence support the concept that angiogenesis is essential for the growth and persistence of solid tumors and their metastases (see, e.g., Folkman, N. Engl. J. Med. 285:1182, 1971; Folkman et al., Nature 339:58, 1989; Kim et al.,  
35        Nature 362:841, 1993; Hori et al., Cancer Res., 51:6180, 1991). Angiogenesis inhibitors are therefore useful for the prevention (e.g., treatment of premalignant conditions), intervention (e.g., treatment of small tumors), and regression (e.g., treatment of large tumors) of cancers (see, e.g., Bergers et al., Science 284:808, 1999).

          There is a need for additional compositions and methods of modulating angiogenesis for the prevention, abrogation, and mitigation of disease.

**B. TWEAK**

The TWEAK protein, which has also been called TREPA and Apo3L, is a member of the tumor necrosis factor (TNF) family and is expressed in a wide variety of human tissues (Chicheportiche et al., J. Biol. Chem., 272(51):32401, 1997; see also Wiley, PCT Publication No. WO 98/35061, 13 August 1998).

5 Like most TNF family members, TWEAK is a Type II membrane protein with an extracellular C-terminal domain. Although TWEAK was originally described as a weak inducer of apoptosis, this induction of cell death was later shown to be indirect (Schneider et al., Eur. J. Immunol. 29:1785, 1999).

Lynch et al. demonstrated that TWEAK directly induces endothelial cell proliferation and angiogenesis (J. Biol. Chem., 274(13):8455, 1999). Picomolar concentrations of recombinant soluble  
10 TWEAK induce proliferation in multiple endothelial cell lines and in aortic smooth-muscle cells, and reduce the requirement for serum and growth factors in culture. Moreover, TWEAK induces a strong angiogenic response in a rat corneal pocket assay. Since TNF family members initiate biological responses by signaling through members of the TNF receptor family, there has been great interest in identifying and characterizing the TWEAK receptor.

15 Marsters et al. reported that TWEAK binds to and signals through a death-domain containing receptor known variously as DR3, Apo3, WSL-1, TRAMP, or LARD (Marsters et al., Current Biology 8(9):525, 1998). Schneider et al., however, showed that TWEAK binds to and signals in Kym-1 cells but that Kym-1 cells do not express the receptor DR3 (Schneider et al., Eur. J. Immunol. 29:1785, 1999). These results suggest the existence of a yet to be identified TWEAK receptor.

20 Because TWEAK induces angiogenesis in vivo, there is a particular need to identify the major functional TWEAK receptor. Once identified, the TWEAK receptor may be used to screen for and develop TWEAK receptor agonists and antagonists for the modulation of angiogenesis and the treatment of human disease.

**SUMMARY OF THE INVENTION**

25 The present invention is based upon the identification and biological characterization of the major functional TWEAK receptor. As described below, cDNA encoding the TWEAK receptor was molecularly cloned from a human endothelial cell expression library.

Although DNA and deduced amino acid sequences corresponding to the TWEAK receptor  
30 identified herein have been reported (see, e.g., Kato et al., PCT Publication No. WO 98/55508, 10 December 1998 and Incyte, PCT Publication No. WO 99/61471, 02 December 1999), it was not heretofore appreciated that these sequences encode a receptor for TWEAK or that the encoded polypeptide is involved in modulating angiogenesis. Similarly, investigators have recently claimed methods of making and using TWEAK receptor antagonists to treat immunological disorders, but without identifying the major TWEAK  
35 receptor or its role in angiogenesis (Rennert, PCT Publication No. WO 00/42073, 20 July 2000). These deficiencies have been addressed, as described herein, by identification of the major TWEAK receptor (TWEAKR) and characterization of its biological activities. The identification of TWEAKR has led to the development of compositions for the modulation of angiogenesis, and also provides screening tools for the identification of diagnostics and therapeutics.

The invention provides methods of modulating angiogenesis in a mammal in need of such treatment comprising administering a therapeutically-effective amount of a composition comprising a TWEAK receptor antagonist or TWEAK receptor agonist. The composition preferably comprises a pharmaceutically acceptable carrier and the mammal is preferably a human.

5 In some more preferred embodiments the composition inhibits angiogenesis and comprises a TWEAK receptor antagonist, such as a soluble TWEAK receptor fragment, an antagonistic antibody, or an antagonist that disrupts the interaction between the TWEAK receptor and a TRAF molecule. In some most preferred embodiments the antagonist comprises amino acids 28-79 of SEQ ID NO:7 or amino acids 28-309 of SEQ ID NO:7. The TWEAK receptor antagonists are preferably used to treat a mammal that has a  
10 disease or condition mediated by angiogenesis, more preferably a disease or condition characterized by ocular neovascularization or a solid tumor. In some embodiments, the mammal is further treated with radiation or with a second chemotherapeutic agent.

In some more preferred embodiments the composition promotes angiogenesis and comprises a TWEAK receptor agonist, such as an agonistic antibody. The TWEAK receptor agonists are preferably used  
15 to treat a vascularization deficiency in cardiac or peripheral tissue, to enhance wound healing or organ transplantation, or in conjunction with bypass surgery or angioplasty.

The invention also provides antagonists comprising a soluble TWEAK receptor fragment for use in medicine, preferably comprising amino acids 28-79 of SEQ ID NO:7 or amino acids 28-309 of SEQ ID NO:7, as well as nucleic acids encoding soluble TWEAK receptor fragments. And the invention provides  
20 for the use of a composition comprising a TWEAK receptor antagonist or TWEAK receptor agonist for the preparation of a medicament for modulating angiogenesis in a mammal in need of such treatment.

The invention further provides methods of identifying a compound that is capable of modulating angiogenesis, including: (a) identifying a test compound that binds to a TWEAK receptor extracellular domain, wherein the test compound is not TWEAK; (b) identifying a test compound that affects the  
25 interaction between a TWEAK and a TWEAK receptor; and (c) identifying a test compound that modulates the interaction between a TWEAK receptor and a TRAF. The invention encompasses compounds identified according to these methods.

The invention also provides a method for targeting a detectable label or chemotherapeutic to vascular tissue comprising contacting the vascular tissue with an antibody that binds TWEAK receptor. In  
30 some preferred embodiments the antibody is conjugated to a radioisotope, chemiluminescent or fluorescent compound, or enzyme. In some preferred embodiments the antibody is conjugated to a cytotoxin.

#### **BRIEF DESCRIPTION OF THE DRAWINGS**

Figure 1 shows a sequence alignment of the human and murine TWEAK receptor polypeptide  
35 sequences. The top sequence is the murine TWEAK receptor polypeptide (SEQ ID NO:5), and the bottom sequence is the human TWEAK receptor polypeptide (SEQ ID NO:4).

Figure 2 shows the effect of TWEAKR-Fc on PMA-induced HRMEC wound closure.

Figure 3 shows the effect of TWEAKR-Fc on EGF-induced HRMEC wound closure.

Figure 4 shows the effect of human TWEAKR-Fc on TWEAK-induced (100 ng/ml) HUVEC proliferation.

Figure 5 shows the effect of human TWEAKR-Fc on FGF-2-induced (10 ng/ml) HUVEC proliferation.

5

### DETAILED DESCRIPTION OF THE INVENTION

The present invention is directed to the TWEAK receptor and methods for identifying and using agonists and antagonists of the TWEAK receptor. The invention provides methods of screening for agonists and antagonists and for treating diseases or conditions mediated by angiogenesis.

10

#### A. Abbreviations and Terminology Used in the Specification

"4-1BB" and "4-1BB ligand" (4-1BB-L) are polypeptides described, inter alia, in U.S. Patent No. 5,674,704, including soluble forms thereof.

"bFGF" is basic fibroblast growth factor.

15

"BSA" is bovine serum albumin.

"CD40 ligand" (CD40L) is a polypeptide described, inter alia, in U.S. Patent No. 5,716,805, including soluble forms thereof.

"CHO" is a Chinese hamster ovary cell line.

"DMEM" is Dulbecco's Modified eagle Medium, a commercially available cell culture medium.

20

"ELISA" is Enzyme-Linked Immunosorbent Assay.

"Flt3L" is Flt3 ligand, a polypeptide described, inter alia, in U.S. Patent No. 5,554,512, including soluble forms thereof.

"HRMEC" are primary human renal microvascular endothelial cells.

"HUVEC" is a line of human umbilical vein endothelial cells.

25

"PBS" is phosphate buffered saline.

"PMA" is phorbol 12-myristate-13-acetate.

"RTKs" are receptor tyrosine kinases.

"Tek," which has also been called Tie2 and ork, is an RTK that is predominantly expressed in vascular endothelium. The molecular cloning of human Tek (ork) has been described by Ziegler, U.S.

30 Patent No. 5,447,860. "Tek antagonists" are described, inter alia, in Cerretti et al., PCT Publication No. WO 00/75323, 14 December 2000.

"TNFR" is a tumor necrosis factor receptor, including soluble forms thereof. "TNFR/Fc" is a tumor necrosis factor receptor-Fc fusion polypeptide.

35 "TRAIL" is TNF-related apoptosis-inducing ligand, a type II transmembrane polypeptide in the TNF family described, inter alia, in U.S. Patent No. 5,763,223, including soluble forms thereof.

"VEGF" is vascular endothelial growth factor, also known as VPF or vascular permeability factor.

#### B. Soluble TWEAK Receptor Polypeptides

As described in the examples below, the native human TWEAK receptor cDNA has the sequence

SEQ ID NO:3, which encodes a 129 residue polypeptide (SEQ ID NO:4). Examination of the DNA sequence predicts a polypeptide having an approximately 78 amino acid extracellular domain (residues 1-78 of SEQ ID NO:4, including the signal peptide), an approximately 23 amino acid transmembrane domain (residues 79-101 of SEQ ID NO:4), and an approximately 28 amino acid intracellular domain (residues 102-129 of SEQ ID NO:4). The TWEAK receptor sequence has also been reported by Kato et al., PCT Publication No. WO 98/55508, 10 December 1998 and by Incyte, PCT Publication No. WO 99/61471, 02 December 1999. As used herein, "TWEAKR" includes polypeptides having these sequences, and in particular comprising amino acids 28-79 of SEQ ID NO:7, as well as naturally occurring variants thereof.

In one aspect of the invention, a soluble TWEAK receptor fragment is used as a TWEAKR antagonist to inhibit angiogenesis and/or to inhibit the binding of TWEAK ligand to TWEAKR.

Soluble polypeptides are capable of being secreted from the cells in which they are expressed. The use of soluble forms of polypeptides is advantageous for certain applications. Purification of the polypeptides from recombinant host cells is facilitated since the polypeptides are secreted, and soluble proteins are generally suited for parenteral administration. A secreted soluble polypeptide may be identified (and distinguished from its non-soluble membrane-bound counterparts) by separating intact cells which express the desired polypeptide from the culture medium, e.g., by centrifugation, and assaying the medium (supernatant) for the presence of the desired polypeptide. The presence of the desired polypeptide in the medium indicates that the polypeptide was secreted from the cells and thus is a soluble form of the polypeptide. Soluble polypeptides may be prepared by any of a number of conventional techniques. A DNA sequence encoding a desired soluble polypeptide may be subcloned into an expression vector for production of the polypeptide, or the desired encoding DNA fragment may be chemically synthesized.

Soluble TWEAKR polypeptides comprise all or part of the TWEAKR extracellular domain, but generally lack the transmembrane domain that would cause retention of the polypeptide at the cell surface. Soluble polypeptides may include part of the transmembrane domain or all or part of the cytoplasmic domain as long as the polypeptide is secreted from the cell in which it is produced. Soluble TWEAKR polypeptides advantageously comprise a native or heterologous signal peptide when initially synthesized to promote secretion from the cell, but the signal sequence is cleaved upon secretion. The term "TWEAKR extracellular domain" is intended to encompass all or part of the native TWEAKR extracellular domain, as well as related forms including but not limited to: (a) fragments, (b) variants, (c) derivatives, and (d) fusion polypeptides. The ability of these related forms to inhibit angiogenesis or other TWEAKR-mediated responses may be determined in vitro or in vivo, using methods such as those exemplified below or using other assays known in the art. Examples of soluble TWEAKR polypeptides are provided below. In some embodiments of the present invention a multimeric form of a soluble TWEAKR polypeptide ("soluble TWEAKR multimer") is used as an antagonist to block the binding of TWEAK to TWEAKR, to inhibit angiogenesis or other TWEAKR-mediated responses.

Soluble TWEAKR multimers are covalently-linked or non-covalently-linked multimers, including dimers, trimers, or higher multimers. Multimers may be linked by disulfide bonds formed between cysteine residues on different soluble TWEAKR polypeptides. One embodiment of the invention is directed to multimers comprising multiple soluble TWEAKR polypeptides joined via covalent or non-covalent

interactions between peptide moieties fused to the soluble TWEAKR polypeptides. Such peptides may be peptide linkers (spacers), or peptides that have the property of promoting multimerization. Leucine zippers and certain polypeptides derived from antibodies are among the peptides that can promote multimerization of soluble TWEAKR polypeptides attached thereto, as described in more detail below. In particular  
5   embodiments, the multimers comprise from two to four soluble TWEAKR polypeptides.

In some embodiments, a soluble TWEAKR multimer is prepared using polypeptides derived from immunoglobulins. Preparation of fusion proteins comprising certain heterologous polypeptides fused to various portions of antibody-derived polypeptides (including the Fc domain) has been described, e.g., by Ashkenazi et al. (Proc. Natl. Acad. Sci. USA 88:10535, 1991); Byrn et al. (Nature 344:677, 1990); and  
10   Hollenbaugh and Aruffo ("Construction of Immunoglobulin Fusion Proteins", in *Current Protocols in Immunology*, Suppl. 4, pages 10.19.1-10.19.11, 1992).

One preferred embodiment of the present invention is directed to a TWEAKR-Fc dimer comprising two fusion proteins created by fusing soluble TWEAKR to an Fc polypeptide. A gene fusion encoding the TWEAKR-Fc fusion protein is inserted into an appropriate expression vector. TWEAKR-Fc fusion proteins  
15   are expressed in host cells transformed with the recombinant expression vector, and allowed to assemble much like antibody molecules, whereupon interchain disulfide bonds form between the Fc moieties to yield divalent soluble TWEAKR. The term "Fc polypeptide" as used herein includes native and mutein forms of polypeptides derived from the Fc region of an antibody. Truncated forms of such polypeptides containing the hinge region that promotes dimerization are also included.

One suitable Fc polypeptide, described in PCT application WO 93/10151, is a single chain polypeptide extending from the N-terminal hinge region to the native C-terminus of the Fc region of a human IgG1 antibody. Another useful Fc polypeptide is the Fc mutein described in U.S. Patent 5,457,035 and by Baum et al., EMBO J. 13:3992, 1994. The amino acid sequence of this mutein is identical to that of the native Fc sequence presented in WO 93/10151, except that amino acid 19 has been changed from Leu to  
25   Ala, amino acid 20 has been changed from Leu to Glu, and amino acid 22 has been changed from Gly to Ala. The mutein exhibits reduced affinity for Fc receptors. Fusion polypeptides comprising Fc moieties, and multimers formed therefrom, offer an advantage of facile purification by affinity chromatography over Protein A or Protein G columns, and Fc fusion polypeptides may provide a longer in vivo half life, which is useful in therapeutic applications, than unmodified polypeptides.

30   In other embodiments, a soluble TWEAKR polypeptide may be substituted for the variable portion of an antibody heavy or light chain. If fusion proteins are made with both heavy and light chains of an antibody, it is possible to form a soluble TWEAKR multimer with as many as four soluble TWEAKR polypeptides.

Alternatively, the soluble TWEAKR multimer is a fusion protein comprising multiple soluble  
35   TWEAKR polypeptides, with or without peptide linkers (spacers), or peptides that have the property of promoting multimerization. Among the suitable peptide linkers are those described in U.S. Patents 4,751,180, 4,935,233, and 5,073,627. A DNA sequence encoding a desired peptide linker may be inserted between, and in the same reading frame as, the DNA sequences encoding TWEAKR, using conventional techniques known in the art. For example, a chemically synthesized oligonucleotide encoding the linker

may be ligated between sequences encoding soluble TWEAKR. In particular embodiments, a fusion protein comprises from two to four soluble TWEAKR polypeptides, separated by peptide linkers.

Another method for preparing soluble TWEAKR multimers involves use of a leucine zipper domain. Leucine zipper domains are peptides that promote multimerization of the proteins in which they are found. Leucine zippers were originally identified in several DNA-binding proteins (Landschulz et al., Science 240:1759, 1988), and have since been found in a variety of different proteins. Among the known leucine zippers are naturally occurring peptides and derivatives thereof that dimerize or trimerize. Examples of leucine zipper domains suitable for producing soluble multimeric proteins are described in PCT application WO 94/10308, and the leucine zipper derived from lung surfactant protein D (SPD) described in Hoppe et al. FEBS Lett. 344:191, 1994. The use of a modified leucine zipper that allows for stable trimerization of a heterologous protein fused thereto is described in Fanslow et al., Semin. Immunol. 6:267, 1994. Recombinant fusion proteins comprising a soluble TWEAKR polypeptide fused to a leucine zipper peptide are expressed in suitable host cells, and the soluble TWEAKR multimer that forms is recovered from the culture supernatant.

For some applications, the soluble TWEAKR multimers of the present invention are believed to provide certain advantages over the use of monomeric forms. Fc fusion polypeptides, for example, typically exhibit an increased in vivo half life as compared to an unmodified polypeptide.

The present invention encompasses the use of various forms of soluble TWEAKR multimers that retain the ability to inhibit angiogenesis or other TWEAKR-mediated responses. The term "soluble TWEAKR multimer" is intended to encompass multimers containing all or part of the native TWEAKR extracellular domain, as well as related forms including, but not limited to, multimers of: (a) fragments, (b) variants, (c) derivatives, and (d) fusion polypeptides of soluble TWEAKR. The ability of these related forms to inhibit angiogenesis or other TWEAKR-mediated responses may be determined in vitro or in vivo, using methods such as those exemplified in the examples or using other assays known in the art.

Among the soluble TWEAKR polypeptides and soluble TWEAKR multimers useful in practicing the present invention are TWEAKR variants that retain the ability to bind ligand and/or inhibit angiogenesis or other TWEAKR-mediated responses. Such TWEAKR variants include polypeptides that are substantially homologous to native TWEAKR, but which have an amino acid sequence different from that of a native TWEAKR because of one or more deletions, insertions or substitutions. Particular embodiments include, but are not limited to, TWEAKR polypeptides that comprise from one to ten deletions, insertions or substitutions of amino acid residues, when compared to a native TWEAKR sequence. Included as variants of TWEAKR polypeptides are those variants that are naturally occurring, such as allelic forms and alternatively spliced forms, as well as variants that have been constructed by modifying the amino acid sequence of a TWEAKR polypeptide or the nucleotide sequence of a nucleic acid encoding a TWEAKR polypeptide.

Generally, substitutions for one or more amino acids present in the native polypeptide should be made conservatively. Examples of conservative substitutions include substitution of amino acids outside of the active domain(s), and substitution of amino acids that do not alter the secondary and/or tertiary structure of TWEAKR. Additional examples include substituting one aliphatic residue for another, such as Ile, Val,

Leu, or Ala for one another, or substitutions of one polar residue for another, such as between Lys and Arg; Glu and Asp; or Gln and Asn, or substitutions of one aromatic residue for another, such as Phe, Trp, or Tyr for one another. Other such conservative substitutions, for example, substitutions of entire regions having similar hydrophobicity characteristics, are known in the art.

5 In some preferred embodiments the TWEAKR variant is at least about 70% identical in amino acid sequence to the amino acid sequence of native TWEAKR; in some preferred embodiments the TWEAKR variant is at least about 80% identical in amino acid sequence to the amino acid sequence of native TWEAKR. In some more preferred embodiments the TWEAKR variant is at least about 90% identical in amino acid sequence to the amino acid sequence of native TWEAKR; in some more preferred embodiments the TWEAKR variant is at least about 95% identical in amino acid sequence to the amino acid sequence of native TWEAKR. In some most preferred embodiments the TWEAKR variant is at least about 98% identical in amino acid sequence to the amino acid sequence of native TWEAKR; in some most preferred embodiments the TWEAKR variant is at least about 99% identical in amino acid sequence to the amino acid sequence of native TWEAKR. Percent identity, in the case of both polypeptides and nucleic acids, may be determined by visual inspection. Percent identity may also be determined using the alignment method of Needleman and Wunsch (J. Mol. Biol. 48:443, 1970) as revised by Smith and Waterman (Adv. Appl. Math 2:482, 1981). Preferably, percent identity is determined by using a computer program, for example, the GAP computer program version 10.x available from the Genetics Computer Group (GCG; Madison, WI, see also Devereux et al., *Nucl. Acids Res.* 12:387, 1984). The preferred default parameters for the GAP program include: (1) a unary comparison matrix (containing a value of 1 for identities and 0 for non-identities) for nucleotides, and the weighted comparison matrix of Gribskov and Burgess, *Nucl. Acids Res.* 14:6745, 1986, as described by Schwartz and Dayhoff, eds., *Atlas of Protein Sequence and Structure*, National Biomedical Research Foundation, pp. 353-358, 1979 for amino acids; (2) a penalty of 30 (amino acids) or 50 (nucleotides) for each gap and an additional 1 (amino acids) or 3 (nucleotides) penalty for each symbol in each gap; (3) no penalty for end gaps; and (4) no maximum penalty for long gaps. Other programs used by one skilled in the art of sequence comparison may also be used. For fragments of TWEAKR, the percent identity is calculated based on that portion of TWEAKR that is present in the fragment.

The present invention further encompasses the use of soluble TWEAKR polypeptides with or without associated native-pattern glycosylation. TWEAKR expressed in yeast or mammalian expression systems (e.g., COS-1 or COS-7 cells) may be similar to or significantly different from a native TWEAKR polypeptide in molecular weight and glycosylation pattern, depending upon the choice of expression system. Expression of TWEAKR polypeptides in bacterial expression systems, such as *E. coli*, provides non-glycosylated molecules. Different host cells may also process polypeptides differentially, resulting in heterogeneous mixtures of polypeptides with variable N- or C-termini.

35 The primary amino acid structure of soluble TWEAKR polypeptides may be modified to create derivatives by forming covalent or aggregative conjugates with other chemical moieties, such as glycosyl groups, lipids, phosphate, acetyl groups and the like. Covalent derivatives of TWEAKR may be prepared by linking particular functional groups to TWEAKR amino acid side chains or at the N-terminus or C-terminus of a TWEAKR polypeptide.

Fusion polypeptides of soluble TWEAKR that are useful in practicing the invention also include covalent or aggregative conjugates of a TWEAKR polypeptide with other polypeptides added to provide novel polyfunctional entities.

5 C. TWEAK Receptor Antibodies

One aspect of the present invention relates to the antigenic epitopes of the TWEAKR extracellular domain. Such epitopes are useful for raising antibodies, and in particular the blocking monoclonal antibodies described in more detail below. Such epitopes or variants thereof can be produced using techniques well known in the art such as solid-phase synthesis, chemical or enzymatic cleavage of a polypeptide, or using recombinant DNA technology.

10 The claimed invention encompasses compositions and uses of antibodies that are immunoreactive with TWEAKR polypeptides. Such antibodies "bind specifically" to TWEAKR polypeptides, meaning that they bind via antigen-binding sites of the antibody as compared to non-specific binding interactions. The terms "antibody" and "antibodies" are used herein in their broadest sense, and include, without limitation, intact monoclonal and polyclonal antibodies as well as fragments such as Fv, Fab, and F(ab')<sub>2</sub> fragments, single-chain antibodies such as scFv, and various chain combinations. The antibodies of the present invention are preferably humanized, and more preferably human. The antibodies may be prepared using a variety of well-known methods including, without limitation, immunization of animals having native or transgenic immune repertoires, phage display, hybridoma and recombinant cell culture, and transgenic plant and animal bioreactors.

20 Both polyclonal and monoclonal antibodies may be prepared by conventional techniques. See, for example, *Monoclonal Antibodies, Hybridomas: A New Dimension in Biological Analyses*, Kennet et al. (eds.), Plenum Press, New York (1980); and *Antibodies: A Laboratory Manual*, Harlow and Land (eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1988).

25 Hybridoma cell lines that produce monoclonal antibodies specific for the polypeptides of the invention are also contemplated herein. Such hybridomas may be produced and identified by conventional techniques. One method for producing such a hybridoma cell line comprises immunizing an animal with a polypeptide, harvesting spleen cells from the immunized animal, fusing said spleen cells to a myeloma cell line, thereby generating hybridoma cells, and identifying a hybridoma cell line that produces a monoclonal antibody that binds the polypeptide. The monoclonal antibodies produced by hybridomas may be recovered by conventional techniques.

30 The monoclonal antibodies of the present invention include chimeric-antibodies, e.g., "humanized" versions of antibodies originally produced in mice or other non-human species. A humanized antibody is an engineered antibody that typically comprises the variable region of a non-human (e.g., murine) antibody, or at least complementarity determining regions (CDRs) thereof, and the remaining immunoglobulin portions derived from a human antibody. Procedures for the production of chimeric and further engineered monoclonal antibodies include those described in Riechmann et al. (*Nature* 332:323, 1988), Liu et al. (*PNAS* 84:3439, 1987), Larrick et al. (*Bio/Technology* 7:934, 1989), and Winter and Harris (*TIPS* 14:139, May,

1993). Such humanized antibodies may be prepared by known techniques and offer the advantage of reduced immunogenicity when the antibodies are administered to humans.

Procedures that have been developed for generating human antibodies in non-human animals may be employed in producing antibodies of the present invention. The antibodies may be partially human or preferably completely human. For example, transgenic mice into which genetic material encoding one or more human immunoglobulin chains has been introduced may be employed. Such mice may be genetically altered in a variety of ways. The genetic manipulation may result in human immunoglobulin polypeptide chains replacing endogenous immunoglobulin chains in at least some, and preferably virtually all, antibodies produced by the animal upon immunization.

Mice in which one or more endogenous immunoglobulin genes have been inactivated by various means have been prepared. Human immunoglobulin genes have been introduced into the mice to replace the inactivated mouse genes. Antibodies produced in the animals incorporate human immunoglobulin polypeptide chains encoded by the human genetic material introduced into the animal. Examples of techniques for the production and use of such transgenic animals to make antibodies (which are sometimes called "transgenic antibodies") are described in U.S. Patent Nos. 5,814,318, 5,569,825, and 5,545,806, which are incorporated by reference herein.

#### D. Inhibitory Antisense, Ribozyme, and Triple Helix Approaches

Modulation of angiogenesis in a tissue or group of cells may also be ameliorated by decreasing the level of TWEAKR gene expression and/or TWEAK receptor-ligand interaction by using TWEAK receptor or ligand gene sequences in conjunction with well-known antisense, gene "knock-out," ribozyme and/or triple helix methods to decrease the level of TWEAK receptor or ligand gene expression. Among the compounds that may exhibit the ability to modulate the activity, expression or synthesis of the TWEAK receptor or ligand gene, including the ability to modulate angiogenesis, are antisense, ribozyme, and triple helix molecules. Such molecules may be designed to reduce or inhibit either unimpaired, or if appropriate, mutant target gene activity. Techniques for the production and use of such molecules are well known to those of skill in the art.

#### E. Recombinant Production of TWEAK Receptor Polypeptides

TWEAKR polypeptides, including soluble TWEAKR polypeptides, fragments, and fusion polypeptides, used in the present invention may be prepared using a recombinant expression system. Host cells transformed with a recombinant expression vector ("recombinant host cells") encoding the TWEAKR polypeptide are cultured under conditions that promote expression of TWEAKR and the TWEAKR is recovered. TWEAKR polypeptides can also be produced in transgenic plants or animals, or by chemical synthesis.

The invention encompasses nucleic acid molecules encoding the TWEAKR polypeptides used in the invention, including: (a) nucleic acids that encode residues 28-79 of SEQ ID NO:7 and fragments thereof that bind TWEAK; (b) nucleic acids that are at least 70%, 80%, 90%, 95%, 98%, or 99% identical to a nucleic acid of (a), and which encode a polypeptide capable of binding TWEAK; and (c) nucleic acids that

hybridize at moderate stringency to a nucleic acid of (a), and which encode a polypeptide capable of binding TWEAK.

Due to degeneracy of the genetic code, there can be considerable variation in nucleotide sequences encoding the same amino acid sequence. Included as embodiments of the invention are nucleic acid sequences capable of hybridizing under moderately stringent conditions (e.g., prewashing solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0) and hybridization conditions of 50°C, 5 X SSC, overnight) to the DNA sequences encoding TWEAKR. The skilled artisan can determine additional combinations of salt and temperature that constitute moderate hybridization stringency (see also, Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1989; Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, 1982; and Ausubel, *Current Protocols in Molecular Biology*, Wiley and Sons, 1989 and later versions, which are incorporated herein by reference). Conditions of higher stringency include higher temperatures for hybridization and post-hybridization washes, and/or lower salt concentration. Percent identity of nucleic acids may be determined using the methods described above for polypeptides, i.e., by methods including visual inspection and the use of computer programs such as GAP.

Any suitable expression system may be employed for the production of recombinant TWEAKR. Recombinant expression vectors include DNA encoding a TWEAKR polypeptide operably linked to suitable transcriptional and translational regulatory nucleotide sequences, such as those derived from a mammalian, microbial, viral, or insect gene. Nucleotide sequences are operably linked when the regulatory sequence functionally relates to the TWEAKR DNA sequence. Thus, a promoter nucleotide sequence is operably linked to a TWEAKR DNA sequence if the promoter nucleotide sequence controls the transcription of the TWEAKR DNA sequence. Examples of regulatory sequences include transcriptional promoters, operators, or enhancers, an mRNA ribosomal binding site, and appropriate sequences which control transcription and translation initiation and termination. A sequence encoding an appropriate signal peptide (native or heterologous) can be incorporated into expression vectors. A DNA sequence for a signal peptide (referred to by a variety of names including secretory leader, leader peptide, or leader) may be fused in frame to the TWEAKR sequence so that the TWEAKR polypeptide is initially translated as a fusion protein comprising the signal peptide. A signal peptide that is functional in the intended host cells promotes extracellular secretion of the TWEAKR polypeptide. The signal peptide is cleaved from the TWEAKR polypeptide upon secretion of TWEAKR from the cell.

Suitable host cells for expression of TWEAKR polypeptides include prokaryotes, yeast and higher eukaryotic cells, including insect and mammalian cells. Appropriate cloning and expression vectors for use with bacterial, fungal, yeast, insect, and mammalian cellular hosts are described, for example, in Pouwels et al. *Cloning Vectors: A Laboratory Manual*, Elsevier, New York, 1985.

Prokaryotes include gram negative or gram positive organisms, for example, *E. coli* or *Bacilli*. Suitable prokaryotic host cells for transformation include, for example, *E. coli*, *Bacillus subtilis*, *Salmonella typhimurium*, and various other species within the genera *Pseudomonas*, *Streptomyces*, and *Staphylococcus*. In a prokaryotic host cell, such as *E. coli*, TWEAKR polypeptides may include an N-terminal methionine

residue to facilitate expression of the recombinant polypeptide in the prokaryotic host cell. The N-terminal Met may be cleaved from the expressed recombinant polypeptide.

Expression vectors for use in prokaryotic host cells generally comprise one or more phenotypic selectable marker gene(s). A phenotypic selectable marker gene is, for example, a gene encoding a protein that confers antibiotic resistance or that supplies an autotrophic requirement. Examples of useful expression vectors for prokaryotic host cells include those derived from commercially available plasmids such as the cloning vector pBR322 (ATCC 37017). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides simple means for identifying transformed cells. An appropriate promoter and a TWEAKR DNA sequence are inserted into the pBR322 vector. Other commercially available vectors include, for example, pKK223-3 (Pharmacia Fine Chemicals, Uppsala, Sweden) and pGEM1 (Promega Biotec, Madison, WI, USA).

Promoter sequences commonly used for recombinant prokaryotic host cell expression vectors include  $\beta$ -lactamase (penicillinase), lactose promoter system (Chang et al., Nature 275:615, 1978; Goeddel et al., Nature 281:544, 1979), tryptophan (trp) promoter system (Goeddel et al., Nucl. Acids Res. 8:4057, 1980; EP-A-36776) and tac promoter (Maniatis, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, p. 412, 1982). A particularly useful prokaryotic host cell expression system employs a phage  $\lambda$  P<sub>L</sub> promoter and a cI857ts thermolabile repressor sequence. Plasmid vectors available from the American Type Culture Collection which incorporate derivatives of the  $\lambda$  P<sub>L</sub> promoter include plasmid pHUB2 (resident in *E. coli* strain JMB9, ATCC 37092) and pPLc28 (resident in *E. coli* RR1, ATCC 53082).

TWEAKR polypeptides may also be expressed in yeast host cells, preferably from the *Saccharomyces* genus (e.g., *S. cerevisiae*). Other genera of yeast, such as *Pichia* or *Kluyveromyces*, may also be employed. Yeast vectors will often contain an origin of replication sequence from a 2 $\mu$  yeast plasmid, an autonomously replicating sequence (ARS), a promoter region, sequences for polyadenylation, sequences for transcription termination, and a selectable marker gene. Suitable promoter sequences for yeast vectors include, among others, promoters for metallothionein, 3-phosphoglycerate kinase (Hitzeman et al., J. Biol. Chem. 255:2073, 1980) or other glycolytic enzymes (Hess et al., J. Adv. Enzyme Reg. 7:149, 1968; Holland et al., Biochem. 17:4900, 1978), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phospho-glucose isomerase, and glucokinase. Other suitable vectors and promoters for use in yeast expression are further described in Hitzeman, EPA-73,657. Another alternative is the glucose-repressible ADH2 promoter described by Russell et al. (J. Biol. Chem. 258:2674, 1982) and Beier et al. (Nature 300:724, 1982). Shuttle vectors replicable in both yeast and *E. coli* may be constructed by inserting DNA sequences from pBR322 for selection and replication in *E. coli* (Amp<sup>r</sup> gene and origin of replication) into the above-described yeast vectors.

The yeast  $\alpha$ -factor leader sequence may be employed to direct secretion of recombinant polypeptides. The  $\alpha$ -factor leader sequence is often inserted between the promoter sequence and the structural gene sequence. See, e.g., Kurjan et al., Cell 30:933, 1982; Bitter et al., Proc. Natl. Acad. Sci. USA 81:5330, 1984. Other leader sequences suitable for facilitating secretion of recombinant polypeptides from

yeast hosts are known to those of skill in the art. A leader sequence may be modified near its 3' end to contain one or more restriction sites. This will facilitate fusion of the leader sequence to the structural gene.

Yeast transformation protocols are known to those of skill in the art. One such protocol is described by Hinnen et al., Proc. Natl. Acad. Sci. USA 75:1929, 1978. The Hinnen et al. protocol selects for Trp<sup>+</sup> transformants in a selective medium, wherein the selective medium consists of 0.67% yeast nitrogen base, 0.5% casamino acids, 2% glucose, 10 µg/ml adenine and 20 µg/ml uracil.

Yeast host cells transformed by vectors containing an ADH2 promoter sequence may be grown for inducing expression in a "rich" medium. An example of a rich medium is one consisting of 1% yeast extract, 2% peptone, and 1% glucose supplemented with 80 µg/ml adenine and 80 µg/ml uracil.

10 Derepression of the ADH2 promoter occurs when glucose is exhausted from the medium.

Insect host cell culture systems also may be employed to express recombinant TWEAKR polypeptides, including soluble TWEAKR polypeptides. Baculovirus systems for production of heterologous polypeptides in insect cells are reviewed by Luckow and Summers, Bio/Technology 6:47, 1988.

15 Mammalian cells are particularly preferred for use as host cells. Examples of suitable mammalian host cell lines include the COS-7 line of monkey kidney cells (ATCC CRL 1651) (Gluzman et al., Cell 23:175, 1981), L cells, C127 cells, 3T3 cells (ATCC CCL 163), Chinese hamster ovary (CHO) cells, HeLa cells, and BHK (ATCC CRL 10) cell lines, and the CV1/EBNA cell line derived from the African green monkey kidney cell line CV1 (ATCC CCL 70) as described by McMahan et al. (EMBO J. 10: 2821, 1991).

20 For the production of therapeutic polypeptides it is particularly advantageous to use a mammalian host cell line which has been adapted to grow in media that does not contain animal proteins.

Established methods for introducing DNA into mammalian cells have been described (Kaufman, R.J., *Large Scale Mammalian Cell Culture*, 1990, pp. 15-69). Additional protocols using commercially available reagents, such as Lipofectamine (Gibco/BRL) or Lipofectamine-Plus, can be used to transfect cells (Felgner et al., Proc. Natl. Acad. Sci. USA 84:7413, 1987). In addition, electroporation can be used to transfect mammalian cells using conventional procedures, such as those in Sambrook et al. *Molecular Cloning: A Laboratory Manual*, 2 ed. Vol. 1-3. Cold Spring Harbor Laboratory Press, 1989). Selection of stable transformants can be performed using methods known in the art, such as, for example, resistance to cytotoxic drugs. Kaufman et al., Meth. in Enzymology 185:487, 1990, describes several selection schemes, such as dihydrofolate reductase (DHFR) resistance. A suitable host strain for DHFR selection can be CHO strain DX-B11, which is deficient in DHFR (Urlaub and Chasin, Proc. Natl. Acad. Sci. USA 77:4216, 1980). A plasmid expressing the DHFR cDNA can be introduced into strain DX-B11, and only cells that contain the plasmid can grow in the appropriate selective media. Other examples of selectable markers that can be incorporated into an expression vector include cDNAs conferring resistance to antibiotics, such as G418 and hygromycin B. Cells harboring the vector can be selected on the basis of resistance to these compounds.

25

30

35

Transcriptional and translational control sequences for mammalian host cell expression vectors can be excised from viral genomes. Commonly used promoter sequences and enhancer sequences are derived from polyoma virus, adenovirus 2, simian virus 40 (SV40), and human cytomegalovirus. DNA sequences

derived from the SV40 viral genome, for example, SV40 origin, early and late promoter, enhancer, splice, and polyadenylation sites can be used to provide other genetic elements for expression of a structural gene sequence in a mammalian host cell. Viral early and late promoters are particularly useful because both are easily obtained from a viral genome as a fragment, which can also contain a viral origin of replication (Fiers et al., Nature 273:113, 1978; Kaufman, Meth. in Enzymology, 1990). Smaller or larger SV40 fragments can also be used, provided the approximately 250 bp sequence extending from the *Hind* III site toward the *Bgl* I site located in the SV40 viral origin of replication site is included.

Additional control sequences shown to improve expression of heterologous genes from mammalian expression vectors include such elements as the expression augmenting sequence element (EASE) derived from CHO cells (Morris et al., *Animal Cell Technology*, 1997, pp. 529-534) and the tripartite leader (TPL) and VA gene RNAs from Adenovirus 2 (Gingeras et al., J. Biol. Chem. 257:13475, 1982). The internal ribosome entry site (IRES) sequences of viral origin allows dicistronic mRNAs to be translated efficiently (Oh and Sarnow, Current Opinion in Genetics and Development 3:295, 1993; Ramesh et al., Nucleic Acids Research 24:2697, 1996). Expression of a heterologous cDNA as part of a dicistronic mRNA followed by the gene for a selectable marker (e.g. DHFR) has been shown to improve transfectability of the host and expression of the heterologous cDNA (Kaufman, Meth. in Enzymology, 1990). Exemplary expression vectors that employ dicistronic mRNAs are pTR-DC/GFP described by Mosser et al., Biotechniques 22:150, 1997, and p2A5I described by Morris et al., *Animal Cell Technology*, 1997, pp. 529-534.

A useful high expression vector, pCAVNOT, has been described by Mosley et al., Cell 59:335, 1989. Other expression vectors for use in mammalian host cells can be constructed as disclosed by Okayama and Berg (Mol. Cell. Biol. 3:280, 1983). A useful system for stable high level expression of mammalian cDNAs in C127 murine mammary epithelial cells can be constructed substantially as described by Cosman et al. (Mol. Immunol. 23:935, 1986). A useful high expression vector, PMLSV N1/N4, described by Cosman et al., Nature 312:768, 1984, has been deposited as ATCC 39890. Additional useful mammalian expression vectors are known in the art.

Regarding signal peptides that may be employed in producing TWEAKR polypeptides, the native TWEAKR signal peptide may be used or it may be replaced by a heterologous signal peptide or leader sequence, if desired. The choice of signal peptide or leader may depend on factors such as the type of host cells in which the recombinant TWEAKR is to be produced. Examples of heterologous signal peptides that are functional in mammalian host cells include the signal sequence for interleukin-7 (IL-7) described in United States Patent 4,965,195, the signal sequence for interleukin-2 receptor described in Cosman et al., Nature 312:768 (1984); the interleukin-4 receptor signal peptide described in EP 367,566; the type I interleukin-1 receptor signal peptide described in U.S. Patent 4,968,607; and the type II interleukin-1 receptor signal peptide described in EP 460,846.

Using the techniques of recombinant DNA including mutagenesis and the polymerase chain reaction (PCR), the skilled artisan can produce DNA sequences that encode TWEAKR polypeptides comprising various additions or substitutions of amino acid residues or sequences, or deletions of terminal or internal residues or sequences, including TWEAKR fragments, variants, derivatives, and fusion polypeptides.

Transgenic animals, including mice, goats, sheep, and pigs, and transgenic plants, including tobacco, tomato, legumes, grasses, and grains, may also be used as bioreactors for the production of TWEAKR polypeptides, including soluble TWEAKR polypeptides. In the case of transgenic animals, it is particularly advantageous to construct a chimeric DNA including a TWEAKR coding sequence operably  
5 linked to cis-acting regulatory sequences that promote expression of the soluble TWEAKR in milk and/or other body fluids (see, e.g., U.S. Patent No. 5,843,705; U.S. Patent No. 5,880,327). In the case of transgenic plants it is particularly advantageous to produce TWEAKR in a particular cell type, tissue, or organ (see, e.g., US Patent No. 5,639,947; U.S. Patent No. 5,889,189).

The skilled artisan will recognize that the procedure for purifying expressed soluble TWEAKR  
10 polypeptides will vary according to the host system employed, and whether or not the recombinant polypeptide is secreted. Soluble TWEAKR polypeptides may be purified using methods known in the art, including one or more concentration, salting-out, ion exchange, hydrophobic interaction, affinity purification, HPLC, or size exclusion chromatography steps. Fusion polypeptides comprising Fc moieties (and multimers formed therefrom) offer the advantage of facile purification by affinity chromatography over  
15 Protein A or Protein G columns.

#### F. Methods of Treatment

Described below are methods and compositions employing the TWEAK receptor or ligand, or the genes encoding the TWEAK receptor or ligand, to promote or suppress angiogenesis in a target tissue or  
20 group of cells. The terms "treat," "treating," "treatment," "therapy," "therapeutic," and the like are intended to include preventative therapy, prophylactic therapy, ameliorative therapy, and curative therapy.

The disclosed polypeptides, compositions, and methods are used to inhibit angiogenesis or other TWEAKR-mediated responses in a mammal in need of such treatment. The term "TWEAKR-mediated response" includes any cellular, physiological, or other biological response that is caused at least in part by  
25 the binding of TWEAK ligand to TWEAKR, or which may be inhibited or suppressed, in whole or in part, by blocking TWEAK from binding to TWEAKR. The treatment is advantageously administered in order to prevent the onset or the recurrence of a disease or condition mediated by angiogenesis, or to treat a mammal that has a disease or condition mediated by angiogenesis. Diseases and conditions mediated by angiogenesis include but are not limited to ocular disorders, malignant and metastatic conditions, and inflammatory  
30 diseases.

Among the ocular disorders that can be treated according to the present invention are eye diseases characterized by ocular neovascularization including, but not limited to, diabetic retinopathy (a major complication of diabetes), retinopathy of prematurity (this devastating eye condition, that frequently leads to chronic vision problems and carries a high risk of blindness, is a severe complication during the care of  
35 premature infants), neovascular glaucoma, retinoblastoma, retrolental fibroplasia, rubeosis, uveitis, macular degeneration, and corneal graft neovascularization. Other eye inflammatory diseases, ocular tumors, and diseases associated with choroidal or iris neovascularization can also be treated according to the present invention.

The present invention can also be used to treat malignant and metastatic conditions such as solid tumors. Solid tumors include both primary and metastatic sarcomas and carcinomas.

The present invention can also be used to treat inflammatory diseases including, but not limited to, arthritis, rheumatism, and psoriasis.

5 Other diseases and conditions that can be treated according to the present invention include benign tumors and preneoplastic conditions, myocardial angiogenesis, hemophilic joints, scleroderma, vascular adhesions, atherosclerotic plaque neovascularization, telangiectasia, and wound granulation.

Disease states that are angiogenic-dependent include coronary or peripheral atherosclerosis and ischemia of any tissue or organ, including the heart, liver, brain, and the like. These types of diseases can be  
10 treated by compositions that promote angiogenesis.

In addition to polypeptides comprising a fragment of TWEAKR extracellular domain, soluble TWEAKR multimers, and antibodies that bind to the TWEAKR extracellular domain, other forms of TWEAKR antagonists can also be administered to achieve a therapeutic effect. Examples of other forms of TWEAKR antagonists include other antibodies such as antibodies against TWEAK, antisense nucleic acids,  
15 ribozymes, muteins, aptamers, and small molecules directed against TWEAKR or against TWEAK.

The methods according to the present invention can be tested in in vivo animal models to confirm the desired prophylactic or therapeutic activity, as well as to determine the optimal therapeutic dosage, prior to administration to humans.

The amount of a particular TWEAKR antagonist that will be effective in a particular method of  
20 treatment depends upon age, type and severity of the condition to be treated, body weight, desired duration of treatment, method of administration, and other parameters. Effective dosages are determined by a physician or other qualified medical professional. Typical effective dosages are about 0.01 mg/kg to about 100 mg/kg body weight. In some preferred embodiments the dosage is about 0.1-50 mg/kg; in some preferred embodiments the dosage is about 0.5-10 mg/kg. The dosage for local administration is typically  
25 lower than for systemic administration. In some embodiments a single administration is sufficient; in some embodiments the TWEAKR antagonist is administered as multiple doses over one or more days.

The TWEAKR antagonists are typically administered in the form of a pharmaceutical composition comprising one or more pharmacologically acceptable carriers. Pharmaceutically acceptable carriers include diluents, fillers, adjuvants, excipients, and vehicles which are pharmaceutically acceptable for the route of  
30 administration, and may be aqueous or oleaginous suspensions formulated using suitable dispersing, wetting, and suspending agents.

Pharmaceutically acceptable carriers are generally sterile and free of pyrogenic agents, and may include water, oils, solvents, salts, sugars and other carbohydrates, emulsifying agents, buffering agents, antimicrobial agents, and chelating agents. The particular pharmaceutically acceptable carrier and the ratio  
35 of active compound to carrier are determined by the solubility and chemical properties of the composition, the mode of administration, and standard pharmaceutical practice.

The compositions as described herein may be contained in a vial, bottle, tube, syringe inhaler or other container for single or multiple administrations. Such containers may be made of glass or a polymer material such as polypropylene, polyethylene, or polyvinylchloride, for example. Preferred containers may

include a seal, or other closure system, such as a rubber stopper that may be penetrated by a needle in order to withdraw a single dose and then re-seal upon removal of the needle. All such containers for injectable liquids, lyophilized formulations, reconstituted lyophilized formulations or reconstitutable powders for injection known in the art or for the administration of aerosolized compositions are contemplated for use in the presently disclosed compositions and methods.

The TWEAKR antagonists are administered to the patient in a manner appropriate to the indication. Thus, for example, a TWEAKR antagonist, or a pharmaceutical composition thereof, may be administered by intravenous, transdermal, intradermal, intraperitoneal, intramuscular, intranasal, epidural, oral, topical, subcutaneous, intracavity, sustained release from implants, peristaltic routes, or by any other suitable technique. Parenteral administration is preferred.

In certain embodiments of the claimed invention, the treatment further comprises treating the mammal with one or more additional chemotherapeutic agents. The additional chemotherapeutic agent(s) may be administered prior to, concurrently with, or following the administration of the TWEAKR antagonist. The use of more than one chemotherapeutic agent is particularly advantageous when the mammal that is being treated has a solid tumor. In some embodiments of the claimed invention, the treatment further comprises treating the mammal with radiation. Radiation, including brachytherapy and teletherapy, may be administered prior to, concurrently with, or following the administration of the second chemotherapeutic agent(s) and/or TWEAKR antagonist.

When the mammal that is being treated has a solid tumor, the method preferably includes the administration of, in addition to a TWEAKR antagonist, one or more chemotherapeutic agents selected from the group consisting of alkylating agents, antimetabolites, vinca alkaloids and other plant-derived chemotherapeutics, nitrosoureas, antitumor antibiotics, antitumor enzymes, topoisomerase inhibitors, platinum analogs, adrenocortical suppressants, hormones, hormone agonists and antagonists, antibodies, immunotherapeutics, blood cell factors, radiotherapeutics, and biological response modifiers.

In some preferred embodiments the method includes administration of, in addition to a TWEAKR antagonist, one or more chemotherapeutic agents selected from the group consisting of cisplatin, cyclophosphamide, mechlorethamine, melphalan, bleomycin, carboplatin, fluorouracil, 5-fluorodeoxyuridine, methotrexate, taxol, asparaginase, vincristine, and vinblastine, lymphokines and cytokines such as interleukins, interferons (including alpha, beta, or delta), and TNF, chlorambucil, busulfan, carmustine, lomustine, semustine, streptozocin, dacarbazine, cytarabine, mercaptopurine, thioguanine, vindesine, etoposide, teniposide, dactinomycin, daunorubicin, doxorubicin, bleomycin, plicamycin, mitomycin, L-asparaginase, hydroxyurea, methylhydrazine, mitotane, tamoxifen, and fluoxymesterone.

In some preferred embodiments the method includes administration of, in addition to a TWEAKR antagonist, one or more chemotherapeutic agents, including various soluble forms thereof, selected from the group consisting of Flt3 ligand, CD40 ligand, interleukin-2, interleukin-12, 4-1BB ligand, anti-4-1BB antibodies, TNF antagonists and TNF receptor antagonists, TRAIL, VEGF antagonists, VEGF receptor (including VEGF-R1 and VEGF-R2, also known as Flt1 and Flk1 or KDR) antagonists, Tek antagonists, and CD148 (also referred to as DEP-1, ECRTF, and PTPRJ, see Takahashi et al., J. Am. Soc. Nephrol. 10:2135-45, 1999) agonists. In some preferred embodiments the TWEAKR antagonists of the invention are used as a

component of, or in combination with, "metronomic therapy," such as that described by Browder et al. and Klement et al. (Cancer Research 60:1878, 2000; J. Clin. Invest. 105(8):R15, 2000; see also Barinaga, Science 288:245, 2000).

5 The polypeptides, compositions, and methods of the present invention may be used as a first line treatment, for the treatment of residual disease following primary therapy, or as an adjunct to other therapies including chemotherapy, surgery, radiation, and other therapeutic methods known in the art.

When the nucleic acid sequences of the present invention are delivered according to the methods disclosed herein, it is advantageous to use a delivery mechanism so that the sequences will be incorporated into a cell for expression. Delivery systems that may advantageously be employed in the contemplated  
10 methods include the use of, for example, viral delivery systems such as retroviral and adenoviral vectors, as well as non-viral delivery systems. Such delivery systems are well known by those skilled in the art.

#### G. Methods of Screening

The TWEAK receptor as described herein may be used in a variety of methods of screening to  
15 isolate, for example, TWEAKR agonists and antagonists. TWEAKR agonists are compounds that promote the biological activity of TWEAKR and TWEAKR antagonists are compounds that inhibit the biological activity of TWEAKR. Compounds identified via the following screening assays can be used in compositions and methods for modulating angiogenesis to treat a variety of disease states. The present invention provides methods of screening for compounds that (1) modulate TWEAK receptor or ligand gene  
20 expression in a target tissue or cell, (2) modulate the TWEAK receptor-ligand interaction to regulate angiogenesis; (3) bind to the TWEAK receptor or ligand to influence angiogenesis; or (4) interfere with or regulate the bound TWEAK receptor-ligand complex's influence on downstream events such as angiogenesis.

The present invention contemplates the use of assays that are designed to identify compounds that  
25 modulate the activity of a TWEAK receptor or ligand gene (i.e., modulate the level of TWEAK gene expression and/or modulate the level of TWEAK gene product activity). Assays may additionally be utilized that identify compounds that bind to TWEAK gene regulatory sequences (e.g., promoter sequences; see e.g., Platt, 1994, J. Biol. Chem. 269, 28558-28562), and that may modulate the level of TWEAK gene expression.

30 Such an assay may involve, for example, the use of a control system, in which transcription and translation of the TWEAK receptor or ligand gene occurs, in comparison to a system including a test compounds suspected of influencing normal transcription or translation of a TWEAK gene. For example, one could determine the rate of TWEAK receptor RNA produced by cardiac cells, and use this to determine if a test compound influences that rate. To assess the influence of a test compound suspected to influence  
35 this normal rate of transcription, one would first determine the rate of TWEAK receptor RNA production in a cardiac cell culture by, for example, Northern Blotting. One could then administer the test compound to a cardiac cell culture under otherwise identical conditions as the control culture. Then the rate of TWEAK receptor RNA in the culture treated with the test compound could be determined by, for example, Northern Blotting, and compared to the rate of TWEAK receptor RNA produced by the control culture cells. An

increase in the TWEAK receptor RNA in the cells contacted with the test compound relative to control cells is indicative of a stimulator of TWEAK receptor gene transcription and/or translation in cardiac cells, while a decrease is indicative of an inhibitor of TWEAK receptor gene transcription and/or translation in cardiac cells.

5           There are a variety of other methods that can be used to determine the level of TWEAK receptor or ligand gene expression as well, and may further be used in assays to determine the influence of a test compound on the level of TWEAK receptor or ligand gene expression. For example, RNA from a cell type or tissue known, or suspected, to express the TWEAK receptor or ligand gene, such as cardiac, may be isolated and tested utilizing hybridization or PCR techniques. The isolated cells can be derived from cell  
10           culture or from a patient. The analysis of cells taken from culture may be a necessary step in the assessment of cells to be used as part of a cell-based gene therapy technique or, alternatively, to test the effect of compounds on the expression of the TWEAK receptor or ligand gene. Such analyses may reveal both quantitative and qualitative aspects of the expression pattern of the TWEAK receptor or ligand gene, including activation or inactivation of TWEAK receptor or ligand gene expression.

15           In one embodiment of such a detection scheme, a cDNA molecule is synthesized from an RNA molecule of interest (e.g., by reverse transcription of the RNA molecule into cDNA). A sequence within the cDNA is then used as the template for a nucleic acid amplification reaction, such as a PCR amplification reaction, or the like. The nucleic acid reagents used as synthesis initiation reagents (e.g., primers) in the reverse transcription and nucleic acid amplification steps of this method are chosen from among the  
20           TWEAK receptor or ligand gene nucleic acid segments described above. The preferred lengths of such nucleic acid reagents are at least 9-30 nucleotides. For detection of the amplified product, the nucleic acid amplification may be performed using radioactively or non-radioactively labeled nucleotides. Alternatively, enough amplified product may be made such that the product may be visualized by standard ethidium bromide staining or by utilizing any other suitable nucleic acid staining method.

25           Additionally, it is possible to perform such TWEAK receptor or ligand gene expression assays "in situ", i.e., directly upon tissue sections (fixed and/or frozen) of patient tissue obtained from biopsies or resections, such that no nucleic acid purification is necessary. TWEAK receptor or ligand gene nucleic acid segments described above can be used as probes and/or primers for such in situ procedures (see, for example, Nuovo, G. J., 1992, "PCR In Situ Hybridization: Protocols And Applications", Raven Press, NY).

30           Compounds identified via assays such as those described herein may be useful, for example, in modulating angiogenesis influenced by the TWEAK receptor-ligand interaction. Such methods of stimulating or inhibiting TWEAK-influenced angiogenesis are discussed herein.

          Alternatively, assay systems may be designed to identify compounds capable of binding the TWEAK receptor or ligand polypeptide of the invention and thereby influencing angiogenesis resulting from  
35           this interaction. Compounds identified may be useful, for example, in modulating the vascularization of target tissues or cells, may be utilized in screens for identifying compounds that disrupt normal TWEAK receptor-ligand interactions, or may in themselves disrupt such interactions.

          The principle of the assays used to identify compounds that bind to the TWEAK receptor or ligand involves preparing a reaction mixture of the TWEAK receptor or ligand and the test compound under

conditions and for a time sufficient to allow the two components to interact and bind, thus forming a complex that can be removed and/or detected in the reaction mixture. These assays can be conducted in a variety of ways. For example, one method to conduct such an assay screening for compounds that bind to the TWEAK receptor, would involve anchoring the TWEAK receptor or the test substance onto a solid phase and detecting TWEAK receptor/test compound complexes anchored on the solid phase at the end of the reaction. In one embodiment of such a method, the TWEAK receptor may be anchored onto a solid surface, and the test compound, which is not anchored, may be labeled, either directly or indirectly. Alternatively, these same methods could be used to screen for test compounds that bind to the TWEAK ligand rather than receptor.

In practice, microtiter plates may conveniently be utilized as the solid phase. The anchored component may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished by simply coating the solid surface with a solution of the protein and drying. Alternatively, an immobilized antibody, preferably a monoclonal antibody, specific for the protein to be immobilized may be used to anchor the protein to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the non-immobilized component is added to the coated surface containing the anchored component. After the reaction is complete, unreacted components are removed (e.g., by washing) under conditions such that any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the previously non-immobilized component is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the previously non-immobilized component is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the previously non-immobilized component (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody).

Alternatively, a reaction can be conducted in a liquid phase, the reaction products separated from unreacted components, and complexes detected; e.g., using an immobilized antibody specific for the TWEAK receptor or ligand or the test compound to anchor any complexes formed in solution, and a labeled antibody specific for the other component of the possible complex to detect anchored complexes.

Those compounds identified as binding agents for either the TWEAK receptor or the TWEAK ligand may further be assessed for their ability to interfere with TWEAK receptor-ligand interaction, as described below, and thereby suppress or promote angiogenesis resulting from TWEAK receptor-ligand interaction. Such compounds may then be used therapeutically to stimulate or inhibit angiogenesis.

The TWEAK receptor and ligand polypeptides of the present invention may also be used in a screening assay to identify compounds and small molecules which specifically interact with the disclosed TWEAK receptor or ligand to either inhibit (antagonize) or enhance (agonize) interaction between these molecules. Thus, for example, polypeptides of the invention may be used to identify antagonists and agonists from cells, cell-free preparations, chemical libraries, and natural product mixtures. The antagonists and agonists may be natural or modified substrates, ligands, enzymes, receptors, *etc.* of the polypeptides of the instant invention, or may be structural or functional mimetics of the polypeptides. Potential antagonists of the TWEAK receptor-ligand interaction of the instant invention may include small molecules, peptides.

and antibodies that bind to and occupy a binding site of the polypeptides, causing them to be unavailable to interact and therefore preventing their normal ability to modulate angiogenesis. Other potential antagonists are antisense molecules which may hybridize to mRNA *in vivo* and block translation of the mRNA into the polypeptides of the instant invention. Potential agonists include small molecules, peptides and antibodies which bind to the instant TWEAK polypeptides and influence angiogenesis as caused by the disclosed interactions of the TWEAK polypeptides of the instant invention.

Small molecule agonists and antagonists are usually less than 10K molecular weight and may possess a number of physiochemical and pharmacological properties that enhance cell penetration, resist degradation and prolong their physiological half-lives. (Gibbs, "Pharmaceutical Research in Molecular Oncology," *Cell*, Vol. 79, (1994).) Antibodies, which include intact molecules as well as fragments such as Fab and F(ab')<sub>2</sub> fragments, may be used to bind to and inhibit the polypeptides of the instant invention by blocking the commencement of a signaling cascade. It is preferable that the antibodies are humanized, and more preferable that the antibodies are human. The antibodies of the present invention may be prepared by any of a variety of well-known methods.

Specific screening methods are known in the art and many are extensively incorporated in high throughput test systems so that large numbers of test compounds can be screened within a short amount of time. The assays can be performed in a variety of formats, including protein-protein binding assays, biochemical screening assays, immunoassays, cell based assays, etc. These assay formats are well known in the art. The screening assays of the present invention are amenable to screening of chemical libraries and are suitable for the identification of small molecule drug candidates, antibodies, peptides and other antagonists and agonists.

One embodiment of a method for identifying molecules which antagonize or inhibit TWEAK receptor-ligand interaction involves adding a candidate molecule to a medium which contains cells that express the polypeptides of the instant invention: changing the conditions of said medium so that, but for the presence of the candidate molecule, the polypeptides would interact; and observing the binding and inhibition of angiogenesis. Binding of the TWEAK receptor and ligand can be determined according to competitive binding assays outlined above, and well known in the art. The angiogenic effect of this binding can be determined via cell proliferation assays such as, for example, cell density assays, or other cell proliferation assays that are also well-known in the art. The activity of the cells contacted with the candidate molecule may then be compared with the identical cells which were not contacted and agonists and antagonists of the TWEAK polypeptide interactions of the instant invention may be identified. The measurement of biological activity may be performed by a number of well-known methods such as measuring the amount of protein present (*e.g.* an ELISA) or of the protein's activity. A decrease in biological stimulation or activation would indicate an antagonist. An increase would indicate an agonist.

Screening assays can further be designed to find molecules that mimic the biological activity resulting from the TWEAK polypeptide interactions of the instant invention. Molecules which mimic the biological activity of a polypeptide may be useful for enhancing the biological activity of the polypeptide. To identify compounds for therapeutically active agents that mimic the biological activity of a polypeptide, it must first be determined whether a candidate molecule binds to the polypeptide. A binding candidate

molecule is then added to a biological assay to determine its biological effects. The biological effects of the candidate molecule are then compared to those of the polypeptide.

Additionally, complex formation within reaction mixtures containing the test compound and normal TWEAK receptor or ligand gene protein may also be compared to complex formation within reaction mixtures containing the test compound and a mutant TWEAK receptor or ligand gene protein. This comparison may be important in those cases wherein it is desirable to identify compounds that disrupt interactions of mutant but not normal TWEAK receptor or ligand gene proteins.

The assay for compounds that interfere with the interaction of the TWEAK receptor or ligand gene products and binding partners can be conducted in a heterogeneous or homogeneous format. Heterogeneous assays involve anchoring either the TWEAK receptor or ligand gene product or the binding partner onto a solid phase and detecting complexes anchored on the solid phase at the end of the reaction. In homogeneous assays, the entire reaction is carried out in a liquid phase. In either approach, the order of addition of reactants can be varied to obtain different information about the compounds being tested. For example, test compounds that interfere with the interaction between the TWEAK receptor or ligand gene products and the binding partners, e.g., by competition, can be identified by conducting the reaction in the presence of the test substance; i.e., by adding the test substance to the reaction mixture prior to or simultaneously with the TWEAK receptor and ligand gene products. Alternatively, test compounds that disrupt preformed complexes, e.g., compounds with higher binding constants that displace one of the components from the complex, can be tested by adding the test compound to the reaction mixture after complexes have been formed. The various formats are described briefly below.

In a heterogeneous assay system, either the TWEAK receptor or ligand gene product, is anchored onto a solid surface, while the non-anchored species is labeled, either directly or indirectly. In practice, microtiter plates are conveniently utilized. The anchored species may be immobilized by non-covalent or covalent attachments. Non-covalent attachment may be accomplished simply by coating the solid surface with a solution of the TWEAK receptor or ligand gene product and drying. Alternatively, an immobilized antibody specific for the species to be anchored may be used to anchor the species to the solid surface. The surfaces may be prepared in advance and stored.

In order to conduct the assay, the partner of the immobilized species is exposed to the coated surface with or without the test compound. After the reaction is complete, unreacted components are removed (e.g., by washing) and any complexes formed will remain immobilized on the solid surface. The detection of complexes anchored on the solid surface can be accomplished in a number of ways. Where the non-immobilized species is pre-labeled, the detection of label immobilized on the surface indicates that complexes were formed. Where the non-immobilized species is not pre-labeled, an indirect label can be used to detect complexes anchored on the surface; e.g., using a labeled antibody specific for the initially non-immobilized species (the antibody, in turn, may be directly labeled or indirectly labeled with a labeled anti-Ig antibody). Depending upon the order of addition of reaction components, test compounds that inhibit complex formation or that disrupt preformed complexes can be detected.

Alternatively, the reaction can be conducted in a liquid phase in the presence or absence of the test compound, the reaction products separated from unreacted components, and complexes detected; e.g., using

an immobilized antibody specific for one of the binding components to anchor any complexes formed in solution, and a labeled antibody specific for the other partner to detect anchored complexes. Again, depending upon the order of addition of reactants to the liquid phase, test compounds that inhibit complex or that disrupt preformed complexes can be identified.

5 In an alternate embodiment of the invention, a homogeneous assay can be used. In this approach, a preformed complex of the TWEAK receptor or ligand gene product is prepared in which either the TWEAK receptor or ligand gene product or its binding partners is labeled, but the signal generated by the label is quenched due to complex formation (see, e.g., U.S. Pat. No. 4,109,496 by Rubenstein which utilizes this approach for immunoassays). The addition of a test substance that competes with and displaces one of the  
10 species from the preformed complex will result in the generation of a signal above background. In this way, test substances that disrupt TWEAK receptor or ligand gene product interaction can be identified.

In a particular embodiment, the TWEAK receptor or ligand gene product can be prepared for immobilization using recombinant DNA techniques. For example, the TWEAK receptor or ligand coding region can be fused to a glutathione-S-transferase (GST) gene using a fusion vector, such as pGEX-5X-1, in  
15 such a manner that its binding activity is maintained in the resulting fusion protein. The interactive binding partner can be purified and used to raise a monoclonal antibody, using methods routinely practiced in the art. This antibody can be labeled with the radioactive isotope  $<^{125}> \text{I}$ , for example, by methods routinely practiced in the art. In a heterogeneous assay, e.g., the GST-TWEAK receptor or ligand fusion protein can be anchored to glutathione-agarose beads. The TWEAK receptor or ligand gene product can then be added  
20 in the presence or absence of the test compound in a manner that allows interaction and binding to occur. At the end of the reaction period, unbound material can be washed away, and the labeled monoclonal antibody can be added to the system and allowed to bind to the complexed components. The interaction between the TWEAK receptor and ligand gene products can be detected by measuring the amount of radioactivity that remains associated with the glutathione-agarose beads. A successful inhibition of the interaction by the test  
25 compound will result in a decrease in measured radioactivity.

Alternatively, a GST-TWEAK receptor gene fusion protein and TWEAK ligand gene product (or *vice versa*) can be mixed together in liquid in the absence of the solid glutathione-agarose beads. The test compound can be added either during or after the species are allowed to interact. This mixture can then be added to the glutathione-agarose beads and unbound material is washed away. Again the extent of inhibition  
30 of the TWEAK receptor-ligand gene product interaction can be detected by adding the labeled antibody and measuring the radioactivity associated with the beads.

In another embodiment of the invention, these same techniques can be employed using peptide fragments that correspond to the binding domains of the TWEAK receptor and/or ligand protein, in place of one or both of the full length proteins. Any number of methods routinely practiced in the art can be used to  
35 identify and isolate the binding sites. These methods include, but are not limited to, mutagenesis of the gene encoding one of the proteins and screening for disruption of binding in a co-immunoprecipitation assay. Compensating mutations in the gene encoding the second species in the complex can then be selected. Sequence analysis of the genes encoding the respective proteins will reveal the mutations that correspond to the region of the protein involved in interactive binding. Alternatively, one protein can be anchored to a

solid surface using methods described in this Section above, and allowed to interact with and bind to its labeled binding partner, which has been treated with a proteolytic enzyme, such as trypsin. After washing, a short, labeled peptide comprising the binding domain may remain associated with the solid material, which can be isolated and identified by amino acid sequencing. Also, once the gene coding for the segments can be engineered to express peptide fragments of the protein, which can then be tested for binding activity and purified or synthesized.

For example, and not by way of limitation, a TWEAK receptor or ligand gene product can be anchored to a solid material as described, above, in this Section by making a GST-TWEAK receptor or ligand fusion protein and allowing it to bind to glutathione agarose beads. The interactive binding partner obtained can be labeled with a radioactive isotope, such as  $^{35}\text{S}$ , and cleaved with a proteolytic enzyme such as trypsin. Cleavage products can then be added to the anchored GST-TWEAK receptor fusion protein or TWEAK ligand fusion protein and allowed to bind. After washing away unbound peptides, labeled bound material, representing the binding partner binding domain, can be eluted, purified, and analyzed for amino acid sequence by well-known methods. Peptides so identified can be produced synthetically or fused to appropriate facilitative proteins using recombinant DNA technology.

The TWEAK receptor-ligand interactions of the invention, *in vivo*, initiate a cascade of events that either stimulate or suppress angiogenesis in a target group of cell or tissue. Molecules, such as nucleic acid molecules, proteins, or small molecules may, in turn, influence this cascade. Compounds that disrupt the TWEAK receptor-ligand interaction effects in this way may be useful in regulating angiogenesis.

The basic principle of the assay systems used to identify compounds that interfere with the angiogenic or anti-angiogenic effect of TWEAK receptor-ligand interaction involves preparing a reaction mixture containing the TWEAK receptor and ligand under conditions and for a time sufficient to allow the two to interact and bind, thus forming a complex. In order to test a compound for inhibitory activity of the effect of this interaction, the reaction mixture is prepared in the presence and absence of the test compound. The test compound may be initially included in the reaction mixture, or may be added at a time subsequent to the addition of the TWEAK receptor-ligand complex. Control reaction mixtures are incubated without the test compound or with a placebo. The inhibition or potentiation of any effect of the TWEAK complex on vascularization is then detected. Normal angiogenic response in the control reaction, but not in the reaction mixture containing the test compound, indicates that the compound interferes with the cascade of events initiated by the TWEAK receptor-ligand interaction. Enhanced angiogenesis in the test compounds-containing culture indicates a stimulator of the TWEAK receptor-ligand complex effect.

### EXAMPLES

The following examples are intended to illustrate particular embodiments and not to limit the scope of the invention.

## EXAMPLE 1

### Identification of the TWEAK Receptor

#### A. Expression Cloning of TWEAK Receptor cDNA

To clone TWEAK Receptor cDNA, an expression vector encoding a growth hormone leader, a leucine zipper multimerization domain, and the C-terminal extracellular domain of human TWEAK (see Chicheportiche et al., J. Biol. Chem. 272(51):32401, 1997) was constructed. This expression vector, which was named pDC409-LZ-TWEAK, comprised the DNA sequence SEQ ID NO:1 and encoded the polypeptide SEQ ID NO:2. pDC409-LZ-TWEAK conditioned supernatants were produced by transient transfection into CV1-EBNA cells. These supernatants were incubated with magnetic beads coated with polyclonal goat anti-mouse antibody that had previously been incubated with a mouse monoclonal antibody against the leucine zipper. Control beads were produced by mixing the coated beads with supernatants from cells transfected with empty vector.

A monolayer of COS cells grown in a T175 flask was transfected with 15 µg of DNA pools of complexity of 100,000 from a HUVEC cDNA expression library. After 2 days these cells were lifted from the flask, and incubated in 1.5 mls of binding media plus 5% non-fat dried milk for 3 hours at 4 degrees C on a rotator wheel. Cells were pre-cleared by adding control beads and rotated at 4 degrees C for an additional 45 minutes after which bead bound cells were removed with a magnet. Pre-clearing was repeated 2-3 times, then TWEAK coated beads were added to the cells and rotated 30 minutes at 4 degrees C. Cells binding the TWEAK beads were separated by use of a magnet and washed 4x in PBS. Plasmid DNA was extracted from these cells by lysing in 0.1% SDS, and electroporating the supernatants in DH101B cells. Colonies were grown overnight on ampicillin selective media. Transformants were pooled and used as a source of plasmid DNA for a further round of panning. After 2 rounds of panning, positive clones were picked from the resulting pool based on their ability to bind TWEAK using a slide binding protocol like that described in Part B, below.

The human TWEAK receptor (also called TWEAKR) cDNA was determined to have the sequence SEQ ID NO:3, which encodes a 129 residue polypeptide (SEQ ID NO:4). Examination of the sequence predicts a polypeptide having an approximately 78 amino acid extracellular domain (residues 1-78 of SEQ ID NO:4, including the signal peptide), an approximately 23 amino acid transmembrane domain (residues 79-101 of SEQ ID NO:4), and an approximately 28 amino acid intracellular domain (residues 102-129 of SEQ ID NO:4). TWEAKR is the smallest known TNF receptor family member. It has a single cysteine-rich repeat region in the extracellular domain, as compared to the 3-4 repeats of other TNF receptor family members. The TWEAKR polypeptide was previously described as a transmembrane protein encoded by a human liver cDNA clone (WO 98/55508, see also WO 99/61471), but had not been identified as the TWEAK receptor. A murine homolog, the FGF-inducible Fn14 (Meighan-Mantha et al., J. Biol. Chem. 274(46):33166, 1999), is approximately 82% identical to the human protein, as shown by the alignment in Figure 1.

The newly identified TWEAK receptor was tested side by side with DR3 (which had been identified as the TWEAK receptor by Marsters et al., Current Biology 8:525, 1998) for the ability to bind to TWEAK.

### B. The TWEAK Receptor Binds to TWEAK

Slides of COS cells were transfected with expression vectors containing TWEAKR, DR3, or vector without insert (control). After two days the cells were incubated with concentrated supernatants from CV-1 cells transfected with a vector encoding the leucine zipper TWEAK extracellular domain fusion protein.

- 5 One hour later the cells were washed and probed with an I-125 labeled antibody against the leucine-zipper domain. The slides were washed, fixed, and autoradiography was performed using x-ray film. The TWEAKR transfected cells bound significant amounts of TWEAK. TWEAK did not bind to the cells transfected with DR3 or the control cells. This experiment confirmed that the TWEAKR polypeptide identified in part A above, rather than DR3, is the major receptor for TWEAK. After discovery of the functional TWEAK receptor, other investigators also reported that DR3 is not the major receptor for TWEAK (Kaptein et al., FEBS Lett., 485(2-3):135, 2000. The TWEAK-TWEAKR binding interaction was further characterized by Scatchard analysis.

- 10 CV-1 cells were transfected with human full length TWEAK and mixed 1:30 with Raji cells, which do not express TWEAK. The cells were incubated with serial dilutions of 125-I labeled human TWEAK receptor-Fc for 2 hours at 4 degrees Celsius. Free and bound probe was separated by microfuging the samples through a phalate oil mixture in plastic tubes. Supernatants and pellets were gamma-counted. Scatchard analyses of TWEAK ligand binding the TWEAK receptor showed a binding affinity constant ( $K_a$ ) of approximately  $4.5 \times 10^8 M^{-1}$ .

### C. The TWEAK Receptor is Strongly Expressed in Cardiac Tissue

- 20 To determine the expression pattern of the TWEAK receptor, Northern blot analyses were performed. Human multiple tissue northern blots were purchased from Clontech (Palo Alto, CA) and probed with P-32 labeled random primed DNA from the TWEAK receptor coding region. The blots were washed and autoradiography was performed using x-ray film. Results showed that in the adult TWEAKR is strongly expressed in heart, placenta, and some skeletal muscle samples. Strong expression in heart tissue further supports the utility of TWEAKR in the diagnosis and treatment of cardiac disease. In contrast to the adult, the fetal tissues expressed TWEAKR more ubiquitously; TWEAKR transcripts were seen in the lung and liver.

30

## **EXAMPLE 2**

### **Preparation of TWEAKR Antagonists and Agonists**

Because TWEAK induces angiogenesis, TWEAKR agonists (such as-agonistic antibodies) may be used to promote angiogenesis and TWEAKR antagonists (such as soluble receptors and antagonistic antibodies) may be used to inhibit angiogenesis.

35

### A. Recombinant Production of Soluble TWEAK Receptor-Fc (TWEAKR-Fc) Fusion Polypeptides

To construct a nucleic acid encoding the TWEAKR extracellular domain fused to Fc, a nucleic acid encoding the N-terminal 79 amino acids from TWEAKR, including the leader (signal peptide), was joined to a nucleic acid encoding an Fc portion from human IgG1. Sequences for this construct are shown as SEQ ID

NO:6 (nucleic acid) and SEQ ID NO:7 (amino acid). In SEQ ID NO:7, residues 1-27 are the predicted signal peptide (predicted to be cleaved upon secretion from the cell: the actual cleavage site was identified by N-terminal sequence analysis, see below), residues 28-79 are from the cysteine-rich TWEAKR extracellular domain, residues 80-81 are from a BglII cloning site, and the remainder is the Fc portion.

5 Upon insertion into a mammalian expression vector, and expression in and secretion from a mammalian host cells, this construct produced a polypeptide designated TWEAKR-Fc. N-terminal sequence analysis determined that the secreted polypeptide designated TWEAKR-Fc had an N-terminus corresponding to residue 28 (Glu) of SEQ ID NO:7. Anti-angiogenic activity of TWEAKR-Fc was demonstrated using assays such as those described in the following examples. An analogous Fc-fusion construct was prepared using  
10 the murine TWEAKR extracellular domain.

#### B. Production of Antibodies that Bind the TWEAKR Extracellular Domain

BALB/c mice are immunized with TWEAKR extracellular domain and spleen cells are collected and used to prepare hybridomas using standard procedures. Hybridoma supernatants are screened, using  
15 ELISA, for the ability to bind TWEAKR. Positives are cloned two times, to insure monoclonality, then isotyped and reassayed for reactivity to TWEAKR. Antibodies and antibody derivatives are also prepared using transgenic mice that express human immunoglobulins and through the use of phage display. The resulting antibodies are tested in assays such as those described in the examples below, to characterize their ability to modulate the TWEAK-TWEAKR interaction, TWEAKR signaling, angiogenesis, and other  
20 downstream biological activities.

Agonistic antibodies are used to promote TWEAK-induced biological activities such as angiogenesis, and antagonistic antibodies are used to inhibit TWEAK-induced biological activities such as angiogenesis. For some applications, the activity of antagonistic antibodies is augmented by conjugation to a radioisotope, to a plant-, fungus-, or bacterial-derived cytotoxin such as ricin A or diphtheria toxin, or to  
25 another chemical poison. And because of the restricted tissue distribution of TWEAKR, antibodies that bind to TWEAKR are particularly useful as targeting agents for imaging or delivering therapeutics to the vasculature. Antibodies that bind TWEAKR can be used, for example, to target a detectable label or chemotherapeutic to the mural cells (pericytes and vascular smooth muscle cells). Detectable labels may include radioisotopes, chemiluminescent and fluorescent compounds, and enzymes. These techniques are  
30 useful, for example, in the diagnosis, staging, and treatment of neoplasms.

#### **EXAMPLE 3**

##### **Activity of TWEAKR-Fc In a Wound Closure Assay**

A planar endothelial cell migration (wound closure) assay was used to quantitate the inhibition of  
35 angiogenesis by TWEAKR-Fc in vitro. In this assay, endothelial cell migration is measured as the rate of closure of a circular wound in a cultured cell monolayer. The rate of wound closure is linear, and is dynamically regulated by agents that stimulate and inhibit angiogenesis in vivo.

Primary human renal microvascular endothelial cells, HRMEC, were isolated, cultured, and used at the third passage after thawing, as described in Martin et al., In Vitro Cell Dev Biol 33:261, 1997. Replicate

circular lesions, "wounds," (600-800 micron diameter) were generated in confluent HRMEC monolayers using a silicon-tipped drill press. At the time of wounding the medium (DMEM + 1% BSA) was supplemented with 20 ng/ml PMA (phorbol-12-myristate-13-acetate), EGF (4 ng/ml), and 0.150 to 5 µg/ml TWEAKR-Fc, or a combination of 40 ng/ml EGF and 0.150 to 5 µg/ml TWEAKR-Fc. The residual wound area was measured as a function of time (0-12 hours) using a microscope and image analysis software (Bioquant, Nashville, TN). The relative migration rate was calculated for each agent and combination of agents by linear regression of residual wound area plotted over time. The results are shown in Figures 2-3.

Compared to huIgG or media+BSA, TWEAKR-Fc inhibited PMA-induced endothelial migration in a dose responsive manner, reducing the rate of migration to unstimulated levels at 5 µg/ml (Figure 2). Neither huIgG nor TWEAKR-Fc inhibited basal (uninduced) migration. When HRMEC migration was induced by EGF, TWEAKR-Fc inhibited endothelial migration in a dose-dependent manner, reducing the rate of migration to unstimulated levels at 5 µg/ml (Figure 3).

#### EXAMPLE 4

##### Activity of TWEAKR-Fc In a Corneal Pocket Assay

A mouse corneal pocket assay was used to quantitate the inhibition of angiogenesis by TWEAKR-Fc in vivo. In this assay, agents to be tested for angiogenic or anti-angiogenic activity are immobilized in a slow release form in a hydron pellet, which is implanted into micropockets created in the corneal epithelium of anesthetized mice. Vascularization is measured as the appearance, density, and extent of vessel ingrowth from the vascularized corneal limbus into the normally avascular cornea.

Hydron pellets, as described in Kenyon et al., Invest Ophthalmol. & Visual Science 37:1625, 1996, incorporated sucralfate with bFGF (90 ng/pellet), bFGF and IgG (14 µg/pellet, control), or bFGF and TWEAKR-Fc (14 µg). The pellets were surgically implanted into corneal stromal micropockets created by micro-dissection 1mm medial to the lateral corneal limbus of 6-8 week old male C57BL mice. After five days, at the peak of neovascular response to bFGF, the corneas were photographed, using a Zeiss slit lamp, at an incipient angle of 35-50° from the polar axis in the meridian containing the pellet. Images were digitized and processed by subtractive color filters (Adobe Photoshop 4.0) to delineate established microvessels by hemoglobin content. Image analysis software (Bioquant, Nashville, TN) was used to calculate the fraction of the corneal image that was vascularized, the vessel density within the vascularized area, and the vessel density within the total cornea.

As shown in Table 1, TWEAKR-Fc (100 pmol) inhibited bFGF (3 pmol)-induced corneal angiogenesis, reducing the vascular density to 50% of that induced by FGF alone or FGF+IgG.

**Table 1**  
**Effect of TWEAKR-Fc on FGF-induced Angiogenesis in the Mouse Corneal Pocket Assay**

| Treatment     | Greater than 50% Reduction in Number and Length of Vessels<br>n/total n (%) |
|---------------|---|
| FGF alone     | 0/2 (0%)  |
| FGF+IgG       | 0/2 (0%)  |
| FGF+TWEAKR-Fc | 6/9 (67%)   |

### EXAMPLE 5

#### Qualitative TRAF Binding to the TWEAK Receptor (TWEAKR) Cytoplasmic Domain

Members of the TRAF family are intra-cellular signaling molecules. Several members of the TRAF family are known to associate with members of the TNF receptor family in order to initiate a signaling cascade that activates the NF-kappa-B pathway, resulting in cell activation and proliferation. A qualitative in vitro binding assay was performed to test whether members of the TRAF family of intra-cellular signaling molecules bind to the cytoplasmic domain of TWEAKR and to learn, therefore, whether the small cytoplasmic domain of TWEAKR is capable of mediating a signal into the cell via the TRAF pathway,

A GST fusion vector consisting of the C-terminal 29 amino acids of TWEAKR fused to glutathione S-transferase was created by sub-cloning the appropriate insert into the pGEX-4T (Amersham Pharmacia Biotech) vector at the BamHI and NotI sites. The product from this vector was expressed in E.coli and bound to sepharose beads as described by Galibert et al., J. Biol. Chem. 273(51):34120, 1998. Similarly constructed beads coated with RANK cytoplasmic domain-GST fusion proteins were used as a positive control, and beads coated with GST alone were used as a negative control. [35S]methionine/cysteine labeled TRAF proteins were produced in reticulocyte lysates (TNT-coupled Reticulocyte Lysate Systems, Promega) according to the manufacturer's protocol. Reticulocyte lysates containing the labeled TRAF molecules were first pre-cleared using the control beads followed incubation with the indicated fusion protein coated beads in binding buffer (50 mM HEPES [pH 7.4], 250 mM NaCl, 0.25% (v/v) Nonidet P-40, 10% glycerol, 2 mM EDTA) at 4 degrees Celsius for 2 hours. After washing 4x with binding buffer bound TRAF molecules eluted from the beads in SDS-loading buffer, separated by SDS-PAGE, dried and exposed to X-ray film.

Binding above background levels was seen with TRAFs 1,2 and 3. No binding above background levels was seen with TRAFs 4,5, and 6. The ability of TWEAKR to bind to TRAFs 1,2, and 3 demonstrates that TWEAKR is capable of inducing a signal to the cell via the TRAF pathway, and therefore transmitting a proliferative signal into the host cell. This experiment provides further evidence that TWEAKR is the functional receptor for TWEAK. It also illustrates a further means by which signaling can be inhibited: by disrupting the TRAF-TWEAKR interaction with a small molecule, or by use of a dominant negative variant of the TRAF molecule.

### EXAMPLE 6

#### Activity of TWEAKR-Fc in an Endothelial Cell Proliferation Assay

An endothelial cell proliferation assay was used to quantitate the inhibition of bFGF or TWEAK induced-proliferation by TWEAKR-Fc in vitro. In this assay, endothelial cell proliferation is measured after 4 days of cell growth in microtiter wells using a cell labeling molecule called calcein AM. Esterases expressed by the cells cleave the calcein and cause it to fluoresce when excited at 485 nm. Uncleaved calcein does not fluoresce. The amount of fluorescence is directly related to the number of endothelial cells in the culture well. Endothelial cell proliferation is often regulated by agents that stimulate and/or inhibit angiogenesis in vivo.

Primary HUVEC (human umbilical vein endothelial cells) were obtained from a commercial source (Clonetics, Walkersville, MD), cultured, and used at passage 2 to 7. Replicate cultures were set up by adding 3000 HUVEC to each microtiter well in endothelial cell basal media (EBM, an endothelial cell basal media that contains no growth factors or serum and is based on the media formulations developed by Dr. Richard Ham at the University of Colorado, Clonetics) plus 0.05% FBS (fetal bovine serum). At the time of culture initiation FGF-2 (fibroblast growth factor-2, 10 ng/ml) or human TWEAK (100 ng/ml) was added to the cultures in the presence of human IgG (huIgG, control) or human TWEAKR-Fc at concentrations ranging from 0.08 µg/ml to 20 µg/ml (0.25 to 20 µg/ml for TWEAK-induced and 0.08 to 6.7 µg/ml for FGF-2-induced). The HUVEC containing cultures were incubated for 4 days at 37 degrees C, 5% CO<sub>2</sub>. On the fourth day of culture 4 µM calcein-AM was added to the cultures and 2 hours later the wells were evaluated for fluorescence. The results, expressed as the average fluorescence (485-530 nm) counts for replicate wells plus or minus the SEM, are shown in Figures 4 and 5.

TWEAKR-Fc specifically inhibited TWEAK-induced HUVEC proliferation in a dose-dependent manner when compared to huIgG which did not effect TWEAK-induced proliferation (Figure 4). In addition, TWEAKR-Fc inhibited the basal proliferation of HUVEC observed during culture in EBM plus 0.05% FBS, as compared to huIgG which did not. Interestingly, TWEAKR-Fc also inhibited FGF-2 mediated HUVEC proliferation at concentrations of greater than 2 µg/ml, as compared to huIgG which did not effect the FGF-2 induced HUVEC proliferative response (Figure 5). These results show that TWEAKR-Fc inhibits HUVEC proliferation induced by the addition of exogenous recombinant human TWEAK. That TWEAKR-Fc partially inhibits serum -induced HUVEC-proliferation indicates HUVEC produce endogenous TWEAK that promotes growth/survival of the EC (endothelial cell) via the TWEAKR. TWEAKR-Fc attenuation of FGF-2 induced proliferation indicates that at least part of the EC response to FGF-2 is dependent on endogenous TWEAK/TWEAKR interaction.

#### EXAMPLE 7

##### Inhibition of Neovascularization by TWEAKR Antagonists in a Murine Cardiac Ischemia/Engraftment Model

Survival of heterotopically transplanted cardiac tissue from one mouse donor to the ear skin of another genetically similar mouse requires adequate neovascularization by the transplanted heart and the surrounding tissue, to promote survival and energy for cardiac muscle function. Inadequate vasculature at the site of transplant causes excessive ischemia to the heart, tissue damage, and failure of the tissue to engraft. Agents that antagonize factors involved in endothelial cell migration and vessel formation can decrease angiogenesis at the site of transplant, thereby limiting graft tissue function and ultimately engraftment itself. A murine heterotopic cardiac isograft model is used to demonstrate the effects of TWEAKR antagonists, including antibodies and TWEAKR-Fc, on neovascularization.

Female BALB/c (=12 weeks of age) recipients are given neonatal heart grafts from donor mice of the same strain. The donor heart tissue is grafted into the left ear pinnae of the recipient on day 0 and the mice are divided into two groups. The control group receives human IgG (Hu IgG) while the other group receives the TWEAKR antagonist, both intraperitoneally. The treatments are continued for five consecutive

days. The functionality of the grafts is determined by monitoring visible pulsatile activity on days 7 and 14 post-engraftment. The inhibition of functional engraftment, as a function of the dose of TWEAKR antagonist, is determined. The histology of the transplanted hearts is examined in order to visualize the effects of the TWEAKR antagonist on edema at the site of transplant and host and donor tissue vasculature (using, e.g., Factor VIII staining).

#### EXAMPLE 6

##### Treatment of Tumors With TWEAKR Antagonists

TWEAKR antagonists, including antibodies and TWEAKR-Fc, are tested in animal models of solid tumors. The effect of the TWEAKR antagonists is determined by measuring tumor frequency and tumor growth.

The relevant disclosures of publications cited herein are specifically incorporated by reference. The examples presented above are not intended to be exhaustive or to limit the scope of the invention. The skilled artisan will understand that variations and modifications and variations are possible in light of the above teachings, and such modifications and variations are intended to be within the scope of the invention.

CLAIMS

I claim:

1. A method of modulating angiogenesis in a mammal in need of such treatment comprising administering a therapeutically-effective amount of a composition comprising a TWEAK receptor antagonist or TWEAK receptor agonist.
2. The method of claim 1 wherein the composition further comprises a pharmaceutically acceptable carrier.
3. The method of claim 1 or claim 2 wherein the mammal is a human.
4. The method of one of claims 1-3 wherein the TWEAK receptor comprises a sequence selected from the group consisting of:
  - (a) amino acids 28-79 of SEQ ID NO:7; and
  - (b) naturally occurring variants of (a).
5. A method of inhibiting angiogenesis according to one of claims 1-4 wherein the composition comprises a TWEAK receptor antagonist.
6. The method of claim 5 wherein the antagonist is selected from the group consisting of soluble receptor fragments, antibodies, antisense and triple helix forming nucleic acids, peptides, and small molecules.
7. The method of claim 6 wherein the antagonist comprises a soluble TWEAK receptor fragment.
8. The method of claim 7 wherein the antagonist further comprises an Fc polypeptide or leucine zipper domain.
9. The method of claim 8 wherein the antagonist comprises an Fc polypeptide fused to: (a) a TWEAK receptor extracellular domain; or (b) a fragment or variant of (a) that is capable of binding TWEAK.
10. The method of claim 9 wherein the TWEAK receptor extracellular domain comprises amino acids 28-79 of SEQ ID NO:7.
11. The method of claim 10 wherein the antagonist comprises amino acids 28-309 of SEQ ID NO:7.
12. The method claim 6 wherein the antagonist comprises an antibody that binds specifically to the TWEAK receptor extracellular domain.
13. The method of claim 12 wherein the antibody is selected from the group consisting of monoclonal antibodies, humanized antibodies, transgenic antibodies, and human antibodies.
14. The method of claim 12 or claim 13 wherein the antibody is conjugated to a radioisotope, to a plant-, fungus-, or bacterial-derived toxin such as ricin A or diphtheria toxin, or to another chemical poison.
15. The method of claim 6 wherein the antagonist disrupts the interaction between the TWEAK receptor and a TRAF molecule.
16. The method of one of claims 5-15 wherein the mammal has a disease or condition mediated by angiogenesis.
17. The method of claim 16 wherein the disease or condition is characterized by ocular neovascularization.

18. The method of claim 16 wherein the disease or condition is a solid tumor.
19. The method of one of claims 16-18 wherein the method further comprises treating the mammal with radiation.
20. The method of one of claims 16-19 wherein the method further comprises treating the mammal with a second chemotherapeutic agent.
21. The method of claim 20 wherein the second chemotherapeutic agent is selected from the group consisting of alkylating agents, antimetabolites, vinca alkaloids and other plant-derived chemotherapeutics, nitrosoureas, antitumor antibiotics, antitumor enzymes, topoisomerase inhibitors, platinum analogs, adrenocortical suppressants, hormones, hormone agonists, hormone antagonists, antibodies, immunotherapeutics, blood cell factors, radiotherapeutics, and biological response modifiers.
22. The method of claim 20 wherein the second chemotherapeutic agent is selected from the group consisting of cisplatin, cyclophosphamide, mechlorethamine, melphalan, bleomycin, carboplatin, fluorouracil, 5-fluorodeoxyuridine, methotrexate, taxol, asparaginase, vincristine, and vinblastine, lymphokines and cytokines such as interleukins, interferons (including alpha, beta, or delta), and TNF, chlorambucil, busulfan, carmustine, lomustine, semustine, streptozocin, dacarbazine, cytarabine, mercaptopurine, thioguanine, vindesine, etoposide, teniposide, dactinomycin, daunorubicin, doxorubicin, bleomycin, plicamycin, mitomycin, L-asparaginase, hydroxyurea, methylhydrazine, mitotane, tamoxifen, and fluoxymesterone.
23. The method of claim 20 wherein the second chemotherapeutic agent is selected from the group consisting of Flt3 ligand, CD40 ligand, interleukin-2, interleukin-12, 4-1BB ligand, anti-4-1BB antibodies, TNF antagonists and TNF receptor antagonists, TRAIL, CD148 agonists, VEGF antagonists, VEGF receptor antagonists, and Tek antagonists.
24. A method of promoting angiogenesis according to one of claims 1-4 wherein the composition comprises a TWEAK receptor agonist.
25. The method of claim 24 wherein the agonist is an agonistic antibody that binds specifically to the TWEAK receptor extracellular domain.
26. The method of claim 25 wherein the antibody is selected from the group consisting of monoclonal antibodies, humanized antibodies, transgenic antibodies, and human antibodies.
27. The method of one of claims 24-26 wherein the agonist is administered:
  - (a) to treat a vascularization deficiency in cardiac or peripheral tissue, including coronary artery disease, myocardial ischemia, myocardial infarction, angina pectoris, peripheral circulation deficits, limb ischemia/ reperfusion injury;
  - (b) to enhance wound healing, organ transplantation, reconnection of severed digits or limbs, or vascular or skin grafting; or
  - (c) in conjunction with bypass surgery or angioplasty.
28. An antagonist comprising a soluble TWEAK receptor fragment for use in medicine.
29. The antagonist of claim 28 further comprising an Fc polypeptide or leucine zipper domain.
30. The antagonist of claim 29 wherein the antagonist comprises an Fc polypeptide fused to: (a) a TWEAK receptor extracellular domain; or (b) a fragment or variant of (a) that is capable of binding TWEAK.

31. The antagonist of claim 30 wherein the TWEAK receptor extracellular domain comprises amino acids 28-79 of SEQ ID NO:7.
32. The antagonist of claim 31 wherein the antagonist comprises amino acids 28-309 of SEQ ID NO:7.
33. A nucleic acid encoding an antagonist according to one of claims 28-32.
34. An expression vector comprising the nucleic acid of claim 33.
35. A recombinant host cell comprising the nucleic acid of claim 33.
36. A method of producing a TWEAK receptor antagonist comprising culturing the host cell of claim 35 under conditions promoting expression of the antagonist.
37. The use of a composition comprising a TWEAK receptor antagonist or TWEAK receptor agonist for the preparation of a medicament for modulating angiogenesis in a mammal in need of such treatment.
38. A method of identifying a compound that is capable of modulating angiogenesis comprising: identifying a test compound that binds to a TWEAK receptor extracellular domain, wherein the test compound is not TWEAK.
39. A method of identifying a compound that is capable of modulating angiogenesis comprising identifying a test compound that affects the interaction between a TWEAK and a TWEAK receptor.
40. A method of identifying a compound that is capable of modulating angiogenesis comprising identifying a test compound that modulates the interaction between a TWEAK receptor and a TRAF.
41. The method of one of claims 38-40 further comprising determining the ability of the test compound to modulate endothelial cell proliferation and/or endothelial cell migration and/or angiogenesis.
42. The method of one of claims 38-41 wherein the modulation is stimulatory.
43. The method of one of claims 38-41 wherein the modulation is inhibitory.
44. A compound identified according to the method of one of claims 38-43, wherein the compound is not TWEAK.
45. A method of modulating the binding of TWEAK to the TWEAK receptor in a mammal in need of such treatment, comprising administering to the mammal an inhibition-effective amount of a composition comprising a TWEAK receptor antagonist selected from the group consisting of : (a) a soluble TWEAK receptor extracellular domain; and (b) an antibody that binds to the TWEAK receptor extracellular domain.
46. A method for targeting a detectable label or chemotherapeutic to vascular tissue comprising contacting vascular tissue with an antibody that binds TWEAK receptor.
47. The method of claim 46 wherein the antibody is conjugated to a radioisotope, chemiluminescent or fluorescent compound, or enzyme.
48. The method of claim 46 wherein the antibody is conjugated to a cytotoxin.

```

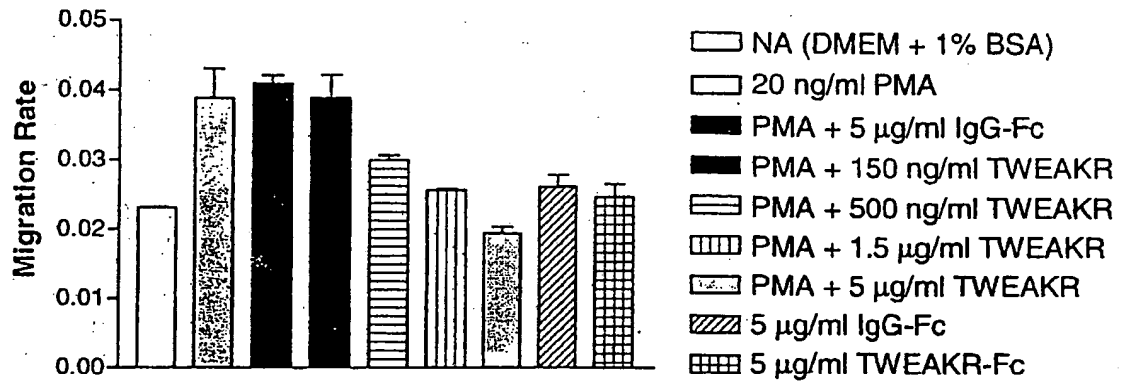
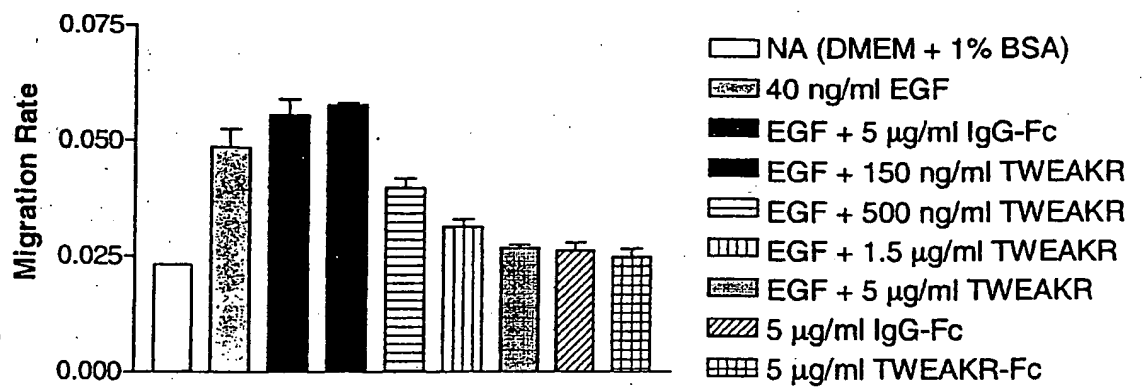
1 MAPGWPRSLPQILVLGFLVLMRAAAGEQA PGTSPCSSGSSWSADLDKCM 50
||| | | .:||| | |:| |||||||.||| |||||||||
1 MARGSLRRLRLLLVLGLWLALLRSVAGEQA PGTAPCSRGSWSADLDKCM 50

51 DCASCPARPHSDFCLGCAAAPPAHFRLLWPILGGALSLVLVALVSSFLV 100
||||| ||||||||| ||||||||| ||| |.| |||
51 DCASCRARPHSDFCLGCAAAPPAFRLLPILGGALSLTFVLGLLSGFLV 100

. . .
101 WRRCCRREKFTTPIEETGGEGC P GVALIQ 129
||||| ||||||||| ||||||| |||||
101 WRRCCRREKFTTPIEETGGEGC P AVALIQ 129

```

Fig. 1

*Fig. 2**Fig. 3*

3/3

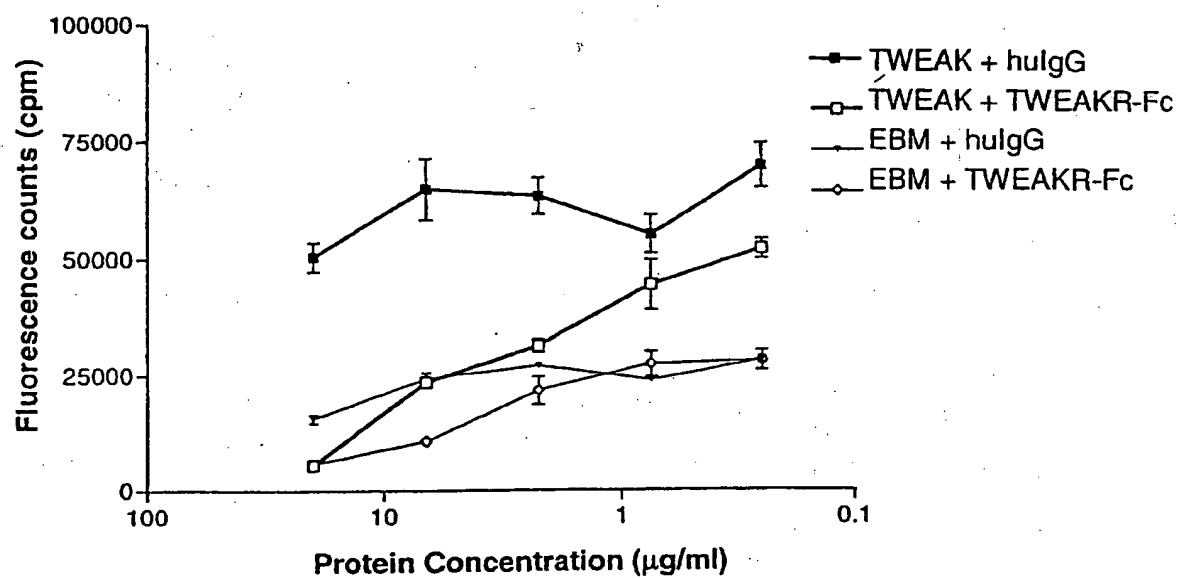


Fig. 4

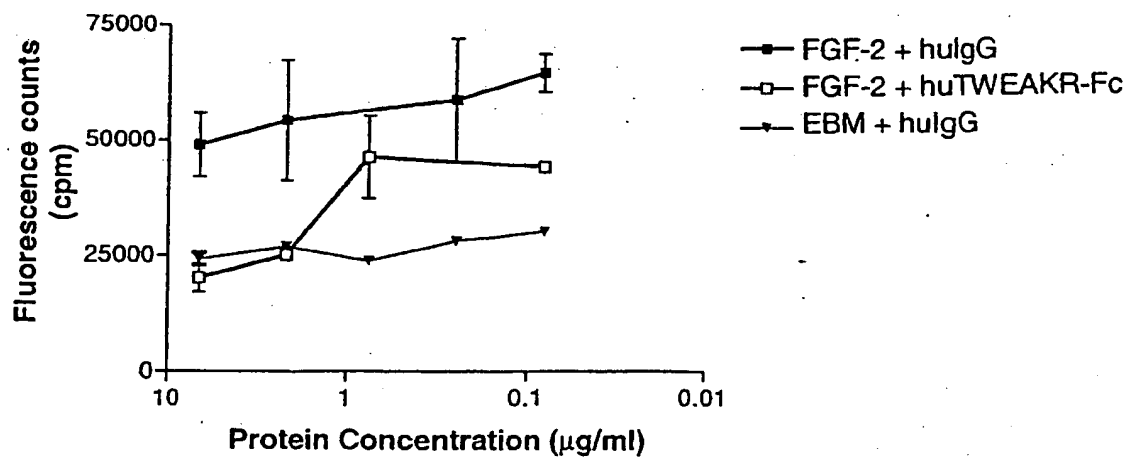


Fig. 5

## SEQUENCE LISTING

<110> WILEY, Steven R.  
IMMUNEX CORPORATION

<120> TWEAK Receptor

<130> 2968-WO

<140> to be assigned

<141> 2000-12-19

<150> 60/172,878

<151> 1999-12-20

<150> 60/203,347

<151> 2000-05-10

<160> 7

<170> PatentIn Ver. 2.0

<210> 1

<211> 898

<212> DNA

<213> Artificial Sequence

<220>

<221> CDS

<222> (52)..(873)

<220>

<223> Description of Artificial Sequence: human TWEAK  
fusion protein construct

<400> 1

```

tctcggagggc cacgcgttta aacgtcgagg tacctatccc gggccgccac c atg gct 57
                                     Met Ala
                                     1

aca ggc tcc cgg acg tcc ctg ctc ctg gct ttt ggc ctg ctc tgc ctg 105
Thr Gly Ser Arg Thr Ser Leu Leu Leu Ala Phe Gly Leu Leu Cys Leu
      5              10              15

ccc tgg ctt caa gag ggc agt gca act agt tct gac cgt atg aaa cag 153
Pro Trp Leu Gln Glu Gly Ser Ala Thr Ser Ser Asp Arg Met Lys Gln
      20              25              30

ata gag gat aag atc gaa gag atc cta agt aag att tat cat ata gag 201
Ile Glu Asp Lys Ile Glu Glu Ile Leu Ser Lys Ile Tyr His Ile Glu
      35              40              45              50

aat gaa atc gcc cgt atc aaa aag ctg att ggc gag cgg act aga tct 249
Asn Glu Ile Ala Arg Ile Lys Lys Leu Ile Gly Glu Arg Thr Arg Ser
              55              60              65

agt ttg ggg agc cgg gca tcg ctg tcc gcc cag gag cct gcc cag gag 297
Ser Leu Gly Ser Arg Ala Ser Leu Ser Ala Gln Glu Pro Ala Gln Glu
              70              75              80

```

```

gag ctg gtg gca gag gag gac cag gac ccg tcg gaa ctg aat ccc cag 345
Glu Leu Val Ala Glu Glu Asp Gln Asp Pro Ser Glu Leu Asn Pro Gln
      85                      90                      95

aca gaa gaa agc cag gat cct gcg cct ttc ctg aac tga cta gtt cgg 393
Thr Glu Glu Ser Gln Asp Pro Ala Pro Phe Leu Asn Arg Leu Val Arg
    100                      105                      110

cct cgc aga agt gca cct aaa ggc cgg aaa aca cgg gct cga aga gcg 441
Pro Arg Arg Ser Ala Pro Lys Gly Arg Lys Thr Arg Ala Arg Arg Ala
    115                      120                      125                      130

atc gca gcc cat tat gaa gtt cat cca cga cct gga cag gac gga gcg 489
Ile Ala Ala His Tyr Glu Val His Pro Arg Pro Gly Gln Asp Gly Ala
      135                      140                      145

cag gca ggt gtg gac ggg aca gtg agt ggc tgg gag gaa gcc aga atc 537
Gln Ala Gly Val Asp Gly Thr Val Ser Gly Trp Glu Glu Ala Arg Ile
      150                      155                      160

aac agc tcc agc cct ctg cgc tac aac cgc cag atc ggg gag ttt ata 585
Asn Ser Ser Ser Pro Leu Arg Tyr Asn Arg Gln Ile Gly Glu Phe Ile
      165                      170                      175

gtc acc cgg gct ggg ctc tac tac ctg tac tgt cag gtg cac ttt gat 633
Val Thr Arg Ala Gly Leu Tyr Tyr Leu Tyr Cys Gln Val His Phe Asp
      180                      185                      190

gag ggg aag gct gtc tac ctg aag ctg gac ttg ctg gtg gat ggt gtg 681
Glu Gly Lys Ala Val Tyr Leu Lys Leu Asp Leu Leu Val Asp Gly Val
    195                      200                      205                      210

ctg gcc ctg cgc tgc ctg gag gaa ttc tca gcc act gcg gcc agt tcc 729
Leu Ala Leu Arg Cys Leu Glu Glu Phe Ser Ala Thr Ala Ala Ser Ser
      215                      220                      225

ctc ggg ccc cag ctc cgc ctc tgc cag gtg tct ggg ctg ttg gcc ctg 777
Leu Gly Pro Gln Leu Arg Leu Cys Gln Val Ser Gly Leu Leu Ala Leu
      230                      235                      240

cgg cca ggg tcc tcc ctg cgg atc cgc acc ctc ccc tgg gcc cat ctc 825
Arg Pro Gly Ser Ser Leu Arg Ile Arg Thr Leu Pro Trp Ala His Leu
      245                      250                      255

aag gct gcc ccc ttc ctc acc tac ttc gga ctc ttc cag gtt cac tga 873
Lys Ala Ala Pro Phe Leu Thr Tyr Phe Gly Leu Phe Gln Val His
      260                      265                      270

gcggccgcgg atctgtttaa actag 898

```

&lt;210&gt; 2

&lt;211&gt; 273

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

<223> Description of Artificial Sequence: human TWEAK  
fusion protein construct

&lt;400&gt; 2

RECTIFIED SHEET (RULE 91)

Met Ala Thr Gly Ser Arg Thr Ser Leu Leu Ala Phe Gly Leu Leu  
 1 5 10 15  
 Cys Leu Pro Trp Leu Gln Glu Gly Ser Ala Thr Ser Ser Asp Arg Met  
 20 25 30  
 Lys Gln Ile Glu Asp Lys Ile Glu Glu Ile Leu Ser Lys Ile Tyr His  
 35 40 45  
 Ile Glu Asn Glu Ile Ala Arg Ile Lys Lys Leu Ile Gly Glu Arg Thr  
 50 55 60  
 Arg Ser Ser Leu Gly Ser Arg Ala Ser Leu Ser Ala Gln Glu Pro Ala  
 65 70 75 80  
 Gln Glu Glu Leu Val Ala Glu Glu Asp Gln Asp Pro Ser Glu Leu Asn  
 85 90 95  
 Pro Gln Thr Glu Glu Ser Gln Asp Pro Ala Pro Phe Leu Asn Arg Leu  
 100 105 110  
 Val Arg Pro Arg Arg Ser Ala Pro Lys Gly Arg Lys Thr Arg Ala Arg  
 115 120 125  
 Arg Ala Ile Ala Ala His Tyr Glu Val His Pro Arg Pro Gly Gln Asp  
 130 135 140  
 Gly Ala Gln Ala Gly Val Asp Gly Thr Val Ser Gly Trp Glu Glu Ala  
 145 150 155 160  
 Arg Ile Asn Ser Ser Ser Pro Leu Arg Tyr Asn Arg Gln Ile Gly Glu  
 165 170 175  
 Phe Ile Val Thr Arg Ala Gly Leu Tyr Tyr Leu Tyr Cys Gln Val His  
 180 185 190  
 Phe Asp Glu Gly Lys Ala Val Tyr Leu Lys Leu Asp Leu Leu Val Asp  
 195 200 205  
 Gly Val Leu Ala Leu Arg Cys Leu Glu Glu Phe Ser Ala Thr Ala Ala  
 210 215 220  
 Ser Ser Leu Gly Pro Gln Leu Arg Leu Cys Gln Val Ser Gly Leu Leu  
 225 230 235 240  
 Ala Leu Arg Pro Gly Ser Ser Leu Arg Ile Arg Thr Leu Pro Trp Ala  
 245 250 255  
 His Leu Lys Ala Ala Pro Phe Leu Thr Tyr Phe Gly Leu Phe Gln Val  
 260 265 270

His

&lt;210&gt; 3

&lt;211&gt; 868

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;220&gt;

&lt;221&gt; CDS

RECTIFIED SHEET (RULE 91)

&lt;222&gt; (53)..(442)

&lt;400&gt; 3

gcttgaattc aataactata acggtcctaa ggtagcgaag aggacgtgca ct atg gct 58  
Met Ala  
1

cgg ggc tgc ctg cgc cgg ttg ctg cgg ctc ctc gtg ctg ggg ctc tgg 106  
Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly Leu Trp  
5 10 15

ctg gcg ttg ctg cgc tcc gtg gcc ggg gag caa gcg cca ggc acc gcc 154  
Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly Thr Ala  
20 25 30

ccc tgc tcc cgc ggc agc tcc tgg agc gcg gac ctg gac aag tgc atg 202  
Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys Cys Met  
35 40 45 50

gac tgc gcg tct tgc agg gcg cga ccg cac agc gac ttc tgc ctg ggc 250  
Asp Cys Ala Ser Cys Arg Ala Arg Pro His Ser Asp Phe Cys Leu Gly  
55 60 65

tgc gct gca gca cct cct gcc ccc ttc cgg ctg ctt tgg ccc atc ctt 298  
Cys Ala Ala Ala Pro Pro Ala Pro Phe Arg Leu Leu Trp Pro Ile Leu  
70 75 80

ggg ggc gct ctg agc ctg acc ttc gtg ctg ggg ctg ctt tct ggc ttt 346  
Gly Gly Ala Leu Ser Leu Thr Phe Val Leu Gly Leu Leu Ser Gly Phe  
85 90 95

ttg gtc tgg aga cga tgc cgc agg aga gag aag ttc acc acc ccc ata 394  
Leu Val Trp Arg Arg Cys Arg Arg Arg Glu Lys Phe Thr Thr Pro Ile  
100 105 110

gag gag acc ggc gga gag ggc tgc cca gct gtg gcg ctg atc cag tga 442  
Glu Glu Thr Gly Gly Glu Gly Cys Pro Ala Val Ala Leu Ile Gln  
115 120 125 130

caatgtgccc cctgccagcc ggggctcgcc cactcatcat tcattcatcc attctagagc 502

cagtctctgc ctcccagacg cggcggggagc caagctcctc caaccacaag ggggggtgggg 562

ggcggtgaaat cacctctgag gcctggggccc agggttcagg ggaaccttcc aagggtgtctg 622

gttgccctgc ctctgggtcc agaacagaaa gggagcctca cgctgggtca cacaaaacag 682

ctgacactga ctaaggaact gcagcatttg cacaggggag gggggtgccc tccttcctag 742

aggccctggg ggccaggctg acttgggggg cagacttgac actaggcccc actcactcag 802

atgtcctgaa attccaccac ggggggtcacc ctgggggggtt agggacctat ttttaacact 862

agagggg 868

&lt;210&gt; 4

&lt;211&gt; 129

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

RECTIFIED SHEET (RULE 91)

&lt;400&gt; 4

Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly  
 1 5 10 15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly  
 20 25 30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys  
 35 40 45

Cys Met Asp Cys Ala Ser Cys Arg Ala Arg Pro His Ser Asp Phe Cys  
 50 55 60

Leu Gly Cys Ala Ala Ala Pro Pro Ala Pro Phe Arg Leu Leu Trp Pro  
 65 70 75 80

Ile Leu Gly Gly Ala Leu Ser Leu Thr Phe Val Leu Gly Leu Leu Ser  
 85 90 95

Gly Phe Leu Val Trp Arg Arg Cys Arg Arg Arg Glu Lys Phe Thr Thr  
 100 105 110

Pro Ile Glu Glu Thr Gly Gly Glu Gly Cys Pro Ala Val Ala Leu Ile  
 115 120 125

Gln

&lt;210&gt; 5

&lt;211&gt; 129

&lt;212&gt; PRT

&lt;213&gt; Mus sp.

&lt;400&gt; 5

Met Ala Pro Gly Trp Pro Arg Ser Leu Pro Gln Ile Leu Val Leu Gly  
 1 5 10 15

Phe Gly Leu Val Leu Met Arg Ala Ala Ala Gly Glu Gln Ala Pro Gly  
 20 25 30

Thr Ser Pro Cys Ser Ser Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys  
 35 40 45

Cys Met Asp Cys Ala Ser Cys Pro Ala Arg Pro His Ser Asp Phe Cys  
 50 55 60

Leu Gly Cys Ala Ala Ala Pro Pro Ala His Phe Arg Leu Leu Trp Pro  
 65 70 75 80

Ile Leu Gly Gly Ala Leu Ser Leu Val Leu Val Leu Ala Leu Val Ser  
 85 90 95

Ser Phe Leu Val Trp Arg Arg Cys Arg Arg Arg Glu Lys Phe Thr Thr  
 100 105 110

Pro Ile Glu Glu Thr Gly Gly Glu Gly Cys Pro Gly Val Ala Leu Ile  
 115 120 125

Gln

RECTIFIED SHEET (RULE 91)

<210> 6  
 <211> 932  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <221> CDS  
 <222> (1)..(930)

<220>  
 <223> Description of Artificial Sequence: human TWEAK  
 receptor fusion protein construct

<400> 6  
 atg gct cgg ggc tgc ctg cgc cgg ttg ctg cgg ctc ctc gtg ctg ggg 48  
 Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly  
 1 5 10 15  
 ctc tgg ctg gcg ttg ctg cgc tcc gtg gcc ggg gag caa gcg cca ggc 96  
 Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly  
 20 25 30  
 acc gcc ccc tgc tcc cgc ggc agc tcc tgg agc gcg gac ctg gac aag 144  
 Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys  
 35 40 45  
 tgc atg gac tgc gcg tct tgc agg gcg cga ccg cac agc gac ttc tgc 192  
 Cys Met Asp Cys Ala Ser Cys Arg Ala Arg Pro His Ser Asp Phe Cys  
 50 55 60  
 ctg ggc tgc gct gca gca cct cct gcc ccc ttc cgg ctg ctt tgg aga 240  
 Leu Gly Cys Ala Ala Ala Pro Pro Ala Pro Phe Arg Leu Leu Trp Arg  
 65 70 75 80  
 tct tgt gac aaa act cac aca tgc cca ccg tgc cca gca cct gaa gcc 288  
 Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Ala  
 85 90 95  
 gag ggc gcg ccg tca gtc ttc ctc ttc ccc cca aaa ccc aag gac acc 336  
 Glu Gly Ala Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr  
 100 105 110  
 ctc atg atc tcc cgg acc cct gag gtc aca tgc gtg gtg gtg gac gtg 384  
 Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val  
 115 120 125  
 agc cac gaa gac cct gag gtc aag ttc aac tgg tac gtg gac ggc gtg 432  
 Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val  
 130 135 140  
 gag gtg cat aat gcc aag aca aag ccg cgg gag gag cag tac aac agc 480  
 Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser  
 145 150 155 160  
 acg tac cgt gtg gtc agc gtc ctc acc gtc ctg cac cag gac tgg ctg 528  
 Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu  
 165 170 175

RECTIFIED SHEET (RULE 91)

```

aat ggc aag gag tac aag tgc aag gtc tcc aac aaa gcc ctc cca gcc 576
Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala
      180                      185                      190

ccc atc gag aaa acc atc tcc aaa gcc aaa ggg cag ccc cga gaa cca 624
Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro
      195                      200                      205

cag gtg tac acc ctg ccc cca tcc cgg gag gag atg acc aag aac cag 672
Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln
      210                      215                      220

gtc agc ctg acc tgc ctg gtc aaa ggc ttc tat ccc agc gac atc gcc 720
Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala
      225                      230                      235                      240

gtg gag tgg gag agc aat ggg cag ccg gag aac aac tac aag acc acg 768
Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr
      245                      250                      255

cct ccc gtg ctg gac tcc gac ggc tcc ttc ttc ctc tat agc aag ctc 816
Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu
      260                      265                      270

acc gtg gac aag agc agg tgg cag cag ggg aac gtc ttc tca tgc tcc 864
Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser
      275                      280                      285

gtg atg cat gag gct ctg cac aac cac tac acg cag aag agc ctc tcc 912
Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser
      290                      295                      300

ctg tct ccg ggt aaa tga ac 932
Leu Ser Pro Gly Lys
305                      310

```

<210> 7  
 <211> 309  
 <212> PRT  
 <213> Artificial Sequence

<220>  
 <223> Description of Artificial Sequence: human TWEAK  
 receptor fusion protein construct

```

<400> 7
Met Ala Arg Gly Ser Leu Arg Arg Leu Leu Arg Leu Leu Val Leu Gly
 1                      5                      10                      15

Leu Trp Leu Ala Leu Leu Arg Ser Val Ala Gly Glu Gln Ala Pro Gly
      20                      25                      30

Thr Ala Pro Cys Ser Arg Gly Ser Ser Trp Ser Ala Asp Leu Asp Lys
      35                      40                      45

Cys Met Asp Cys Ala Ser Cys Arg Ala Arg Pro His Ser Asp Phe Cys
      50                      55                      60

Leu Gly Cys Ala Ala Ala Pro Pro Ala Pro Phe Arg Leu Leu Trp Arg
      65                      70                      75                      80

```

RECTIFIED SHEET (RULE 91)

Ser Cys Asp Lys Thr His Thr Cys Pro Pro Cys Pro Ala Pro Glu Ala  
 85 90 95  
 Glu Gly Ala Pro Ser Val Phe Leu Phe Pro Pro Lys Pro Lys Asp Thr  
 100 105 110  
 Leu Met Ile Ser Arg Thr Pro Glu Val Thr Cys Val Val Val Asp Val  
 115 120 125  
 Ser His Glu Asp Pro Glu Val Lys Phe Asn Trp Tyr Val Asp Gly Val  
 130 135 140  
 Glu Val His Asn Ala Lys Thr Lys Pro Arg Glu Glu Gln Tyr Asn Ser  
 145 150 155 160  
 Thr Tyr Arg Val Val Ser Val Leu Thr Val Leu His Gln Asp Trp Leu  
 165 170 175  
 Asn Gly Lys Glu Tyr Lys Cys Lys Val Ser Asn Lys Ala Leu Pro Ala  
 180 185 190  
 Pro Ile Glu Lys Thr Ile Ser Lys Ala Lys Gly Gln Pro Arg Glu Pro  
 195 200 205  
 Gln Val Tyr Thr Leu Pro Pro Ser Arg Glu Glu Met Thr Lys Asn Gln  
 210 215 220  
 Val Ser Leu Thr Cys Leu Val Lys Gly Phe Tyr Pro Ser Asp Ile Ala  
 225 230 235 240  
 Val Glu Trp Glu Ser Asn Gly Gln Pro Glu Asn Asn Tyr Lys Thr Thr  
 245 250 255  
 Pro Pro Val Leu Asp Ser Asp Gly Ser Phe Phe Leu Tyr Ser Lys Leu  
 260 265 270  
 Thr Val Asp Lys Ser Arg Trp Gln Gln Gly Asn Val Phe Ser Cys Ser  
 275 280 285  
 Val Met His Glu Ala Leu His Asn His Tyr Thr Gln Lys Ser Leu Ser  
 290 295 300  
 Leu Ser Pro Gly Lys  
 305

RECTIFIED SHEET (RULE 91)

# The Fn14 Immediate-Early Response Gene Is Induced During Liver Regeneration and Highly Expressed in Both Human and Murine Hepatocellular Carcinomas

Sheau-Line Y. Feng,\* Yan Guo,\*  
Valentina M. Factor,<sup>†</sup> Snorri S. Thorgeirsson,<sup>†</sup>  
Daphne W. Bell,<sup>‡</sup> Joseph R. Testa,<sup>‡</sup>  
Kimberly A. Peifley,\* and Jeffrey A. Winkles\*<sup>§</sup>

From the Department of Vascular Biology,\* Holland Laboratory, American Red Cross, Rockville, Maryland; the Department of Biochemistry and Molecular Biology and the Institute for Biomedical Sciences,<sup>†</sup> George Washington University Medical Center, Washington, D.C.; the Laboratory of Experimental Carcinogenesis,<sup>‡</sup> Division of Basic Sciences, National Cancer Institute, Bethesda, Maryland; and the Human Genetics Program,<sup>§</sup> Fox Chase Cancer Center, Philadelphia, Pennsylvania

Polypeptide growth factors stimulate mammalian cell proliferation by binding to specific cell surface receptors. This interaction triggers numerous biochemical responses including the activation of protein phosphorylation cascades and the enhanced expression of specific genes. We have identified several fibroblast growth factor (FGF)-inducible genes in murine NIH 3T3 cells and recently reported that one of them, the FGF-inducible 14 (Fn14) immediate-early response gene, is predicted to encode a novel, cell surface-localized type Ia transmembrane protein. Here, we report that the human Fn14 homolog is located on chromosome 16p13.3 and encodes a 129-amino acid protein with ~82% sequence identity to the murine protein. The human Fn14 gene, like the murine Fn14 gene, is expressed at elevated levels after FGF, calf serum or phorbol ester treatment of fibroblasts *in vitro* and is expressed at relatively high levels in heart and kidney *in vivo*. We also report that the human Fn14 gene is expressed at relatively low levels in normal liver tissue but at high levels in liver cancer cell lines and in hepatocellular carcinoma specimens. Furthermore, the murine Fn14 gene is rapidly induced during liver regeneration *in vivo* and is expressed at high levels in the hepatocellular carcinoma nodules that develop in the *c-myc*/transforming growth factor- $\alpha$ -driven and the hepatitis B virus X protein-driven transgenic mouse models of hepatocarcinogenesis. These results indicate that Fn14 may play a role in hepatocyte growth control and liver neoplasia. (*Am J Pathol* 2000; 156:1253-1261)

Polypeptide mitogens such as fibroblast growth factor (FGF)-1 and platelet-derived growth factor-BB stimulate cell cycle progression by binding to specific receptor tyrosine kinases and thereby activating intracellular signal transduction pathways.<sup>1</sup> The activation of cytoplasmic signaling molecules promotes changes in gene expression that are critical for the cellular growth response. Numerous growth factor- and/or serum-inducible genes have been identified and classified into one of three groups: immediate-early, delayed-early, or late response genes.<sup>2</sup> Immediate-early response genes are rapidly and transiently expressed following mitogenic stimulation of quiescent cells and their transcriptional activation does not require *de novo* protein synthesis. Delayed-early response genes are first expressed a few hours later, in the early to middle portions of the G1 phase, and transcript levels often remain elevated for the remainder of the cell cycle. Late response genes are generally expressed only during the S phase of the cell cycle. Both delayed-early and late response genes require *de novo* protein synthesis for their transcriptional activation. Growth factor-inducible genes encode many types of proteins, including transcription factors, cell cycle regulators, extracellular matrix proteins and metabolic enzymes.<sup>2-4</sup>

Several years ago our laboratory used a differential display approach to isolate cDNA fragments representing FGF-1-inducible genes in murine NIH 3T3 fibroblasts.<sup>5,6</sup> One of the immediate-early response genes presently under investigation, the FGF-inducible 14 (Fn14) gene, is located on mouse chromosome 17 and is predicted to encode a 129-amino acid (aa) type Ia transmembrane protein with no significant sequence similarity to any known protein.<sup>7</sup> Furthermore, we have shown that Fn14 is localized on the plasma membrane and that constitutive Fn14 expression in transfected NIH 3T3 fibroblasts decreases cellular adhesion to extracellular matrix proteins and inhibits growth and migration *in vitro*.<sup>7</sup>

Supported in part by National Institutes of Health grants HL-39727 (to J. A. W.) and CA-06927 (to Fox Chase Cancer Center) and by an appropriation from the Commonwealth of Pennsylvania.

Accepted for publication December 21, 1999.

Address reprint requests to Jeffrey A. Winkles, Department of Vascular Biology, Holland Laboratory, American Red Cross, 15601 Crabbs Branch Way, Rockville, MD 20855. E-mail: winkles@usa.redcross.org.

In this paper, we report that the human Fn14 gene encodes a protein with ~82% amino acid sequence identity to the murine Fn14 protein. This gene is located on chromosome 16 and, like its murine homolog, it is activated following growth factor, serum or phorbol ester treatment of quiescent fibroblasts. Additionally, we show that the human Fn14 gene is expressed at relatively high levels in hepatocellular carcinoma (HCC) specimens. We also report that the Fn14 gene is rapidly induced during liver regeneration in the mouse and activated in two different transgenic mouse models of hepatocarcinogenesis.

## Materials and Methods

### Cell Culture

Murine NIH 3T3 fibroblasts were obtained from the American Type Culture Collection (ATCC, Manassas, VA) and grown as described.<sup>7</sup> Human M426 embryonic lung fibroblasts (kind gift of Dr. J. Rubin, National Institutes of Health) were grown at 37°C in Dulbecco's modified Eagle's medium (DMEM; Mediatech, Herndon, VA) supplemented with 10% fetal bovine serum (FBS; Hyclone, Logan, UT), 2 mmol/L glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin, and 0.25 µg/ml amphotericin B (Mediatech). They were incubated in DMEM/Ham's F-12 medium (50/50 mix) supplemented with 5 µg/ml insulin, 5 µg/ml transferrin, 5 ng/ml selenious acid (Collaborative Biomedical Products, Bedford, MA) for 48 hours to induce cellular quiescence. The cells were then either left untreated or treated for various time periods with one of the following: 10 ng/ml human recombinant FGF-1 (kind gift of Dr. W. Burgess; Holland Laboratory, Rockville, MD) in combination with 5 units/ml heparin (Upjohn, Kalamazoo, MI), 10% FBS or 30 ng/ml phorbol myristate acetate (PMA; Sigma, St. Louis, MO). The human liver cell lines were obtained from either the ATCC, the Qidong Liver Cancer Institute, or Dr. C. Harris (National Institutes of Health) and grown according to the provider's instructions.

### Human Fn14 cDNA Sequence Analysis

Homologous sequences to the murine Fn14 cDNA nucleotide sequence were identified using the National Center for Biotechnology Information BLAST program to search the GenBank human expressed sequence tag (EST) database. Several EST clones with a high degree of sequence identity were found. Two clones were obtained from the IMAGE Consortium through Lawrence Livermore National Laboratory and one of these (GenBank accession no. T57612) was sequenced in its entirety. Sequencing was done either automatically using an Applied Biosystems model 373A DNA sequencer and a Dye Terminator Cycle Sequencing kit (Perkin Elmer, Foster City, CA) or manually using a Sequenase 2.0 kit (U.S. Biochemical) and [ $\alpha$ -<sup>35</sup>S]dATP (1000 Ci/mmol, Amersham, Cleveland, OH). The predicted human Fn14 protein sequence was analyzed using several programs

(SignalP, ScanProsite, PSORT II, TMpred, Piscataway, NJ) accessed through the ExPASy Molecular Biology Server. The nucleotide and deduced amino acid sequence reported in this paper has been deposited in the GenBank database under accession no. AF191148.

### RNA Isolation and Northern Blot Hybridization

Total RNA was isolated from NIH 3T3 and M426 cells using RNA Stat-60 (Tel-Test, Friendswood, TX) according to the manufacturer's instructions. Poly(A)<sup>+</sup> RNA was isolated from human liver cell lines, regenerating mouse liver and liver tissue harvested from *c-myc*/transforming growth factor (TGF)- $\alpha$  double transgenic mice<sup>8,9</sup> using the guanidium/cesium chloride method and oligo d(T)-cellulose chromatography as described.<sup>10</sup> RNA samples (10 µg of total RNA or 2 µg of poly(A)<sup>+</sup> RNA) were denatured and subjected to electrophoresis in 1.2% agarose gels containing 2.2 mol/L formaldehyde. The gels were stained with ethidium bromide to verify that each lane contained similar amounts of undegraded rRNA. RNA was transferred onto Zetabind nylon membranes (Cuno Inc, Meriden, CT) by electroblotting and cross-linked to the membrane by UV light irradiation using a Stratalinker (Stratagene, La Jolla, CA). Several Northern blots were purchased from commercial sources. A blot containing 2 µg of poly(A)<sup>+</sup> RNA isolated from various human tissues was obtained from Clontech, Palo Alto, CA. A blot containing 20 µg of total RNA isolated from several different human tumors including a HCC specimen (catalog no. D3100-01) and a blot containing 20 µg of total RNA isolated from three HCC specimens and one cholangiocellular carcinoma specimen (catalog no. D5080-01) were purchased from Invitrogen (Carlsbad, CA). Membrane prehybridization, hybridization, and washing conditions were as described.<sup>7</sup> The cDNA hybridization probes were: (a) human Fn14, ~1.0-kb *Eco*RI/*Xho*I fragment of pBluescript/hFn14; (b) mouse Fn14, ~1.0-kb *Eco*RI/*Xho*I fragment of pBluescript/mFn14; (c) mouse  $\alpha$ -actin, ~1.1-kb *Eco*RI fragment of pVAA (kind gift of Dr. G. Liao, Genetic Therapy Inc.); and (d) rat albumin, ~1.0-kb *Pst*I fragment of pRSA13 (kind gift of Dr. T. Sargent, National Institutes of Health). The probes were radiolabeled with [ $\alpha$ -<sup>32</sup>P]dCTP as described.<sup>7</sup>

### Chromosomal Mapping by Fluorescence in Situ Hybridization (FISH)

Normal human metaphase spreads were prepared according to the method of Fan et al.<sup>11</sup> Human lymphocytes were cultured for 72 hours at 37°C in RPMI 1640 medium containing phytohemagglutinin and 10% FBS. Cultures were synchronized by treatment with 5-bromodeoxyuridine (0.18 mg/ml, Sigma) for 16 hours, followed by release from the block by incubation in fresh medium containing thymidine (2.5 µg/ml) for 6 hours. Metaphase cells were harvested and chromosome spreads were prepared according to standard procedures. FISH and detection of immunofluorescence were performed essentially as described previously.<sup>12</sup> Briefly, pBluescript/

hFn14 plasmid DNA (1 µg) was biotinylated in a nick translation reaction containing 10 µmol/L biotin-16-dUTP (Boehringer Mannheim, Indianapolis, IN) and 2 units DNA polymerase I/DNase I (Gibco BRL, Rockville, MD). Slides were treated with RNase (100 µg/ml in 2× standard saline citrate, SSC) for 1 hour at 37°C, rinsed in 2× SSC, dehydrated in a cold ethanol series, and hybridized overnight at 37°C. They were then washed twice in 50% formamide in 2× SSC at 43°C for 10 minutes, twice in 2× SSC at 37°C for 4 minutes, and once in 4× SSC/0.05% Tween 20 at room temperature for 5 minutes. Slides were removed from the buffer, and then the hybridized probe was detected with fluorescein-labeled avidin (Oncor, Gaithersburg, MD). Signals were amplified by adding a layer of anti-avidin antibody (Oncor), followed by a second layer of fluorescein-labeled avidin according to the manufacturer's instructions. The chromosome preparations were stained with diaminidino-2-phenylindole (DAPI) and observed using a Zeiss Axiophot fluorescence microscope. Digitized images were captured with a cooled CCD camera connected to a computer work station. Images of DAPI staining and fluorescein signals were merged using Oncor Image software, version 1.6.

#### Partial Hepatectomy

C57Bl6 × CBA FI hybrid mice, 7 weeks old, were subjected to a standard 70% partial hepatectomy (PH) as described.<sup>10</sup> They were then sacrificed after various time periods in groups of three, and remnant livers were harvested and pooled for RNA isolation as described above.

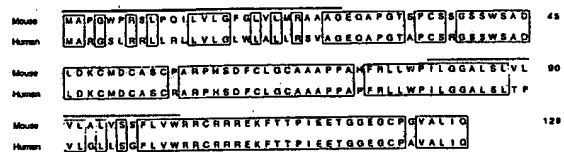
#### In Situ Hybridization

Serial sections of liver tissue from hepatitis B virus X protein (HBx) transgenic mice were the kind gift of Dr. G. Jay (OriGene Technologies). *In situ* hybridization analysis using sense and antisense murine Fn14 riboprobes was performed as described previously.<sup>7</sup>

### Results

#### Human Fn14 cDNA Sequence Analysis

First, we obtained and sequenced a human Fn14 cDNA clone. The BLAST program was used to search the human EST database with the murine Fn14 cDNA nucleotide sequence and several cDNAs were identified. A clone from a human placenta cDNA library was obtained from the IMAGE Consortium and both strands of the ~990-bp insert were sequenced. The DNA sequence, which contained a 29-nucleotide (nt) 5'-untranslated region, a 387-nt open-reading frame, and a 570-nt 3'-untranslated region with a polyadenylation signal and poly(A) tract, had ~80% overall sequence identity to the murine Fn14 cDNA sequence. The open-reading frame encoded a 129-aa protein with a molecular mass of 13,911 daltons and an isoelectric point of 9.37. This protein, like its murine homolog,<sup>7</sup> is predicted to contain a 27-aa signal peptide sequence, a 53-aa extracellular



**Figure 1.** Comparison of the mouse and human Fn14 deduced amino acid sequences. Identical residues are boxed and the numbers to the right refer to the last amino acids on the lines. The solid line indicates the predicted signal peptide sequence and the dotted line indicates the predicted transmembrane domain.

domain, a 21-aa membrane-anchoring domain, and a 28-aa cytoplasmic domain (Figure 1). Human Fn14 has ~82% amino acid sequence identity to murine Fn14 if the signal peptide sequences are included in the analysis. The mature 102-aa human and murine Fn14 proteins have ~90% amino acid sequence identity.

#### Chromosomal Location of the Fn14 Gene

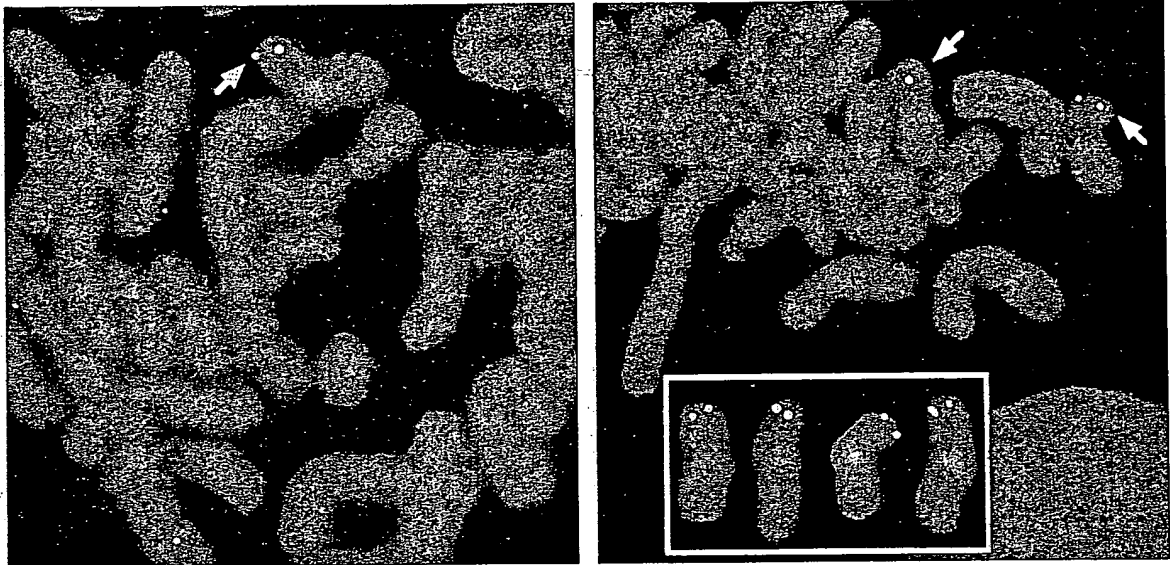
We next determined the chromosomal position of the human Fn14 gene. The Fn14 locus was mapped to chromosome 16p13.3 by FISH. In an analysis of 23 metaphase spreads, 34% (49/145) of all fluorescence signals hybridized to chromosome 16p. All chromosome-specific signals were localized to 16p13.3 (Figure 2). At least one signal specific for chromosome 16 was observed in 21 of the 23 metaphase spreads examined. The distribution of chromosome 16 signals was as follows: one chromatid (1 cell), two chromatids (14 cells), three chromatids (4 cells), four chromatids (2 cells).

#### Regulation of Fn14 mRNA Expression in M426 Cells

The murine Fn14 gene is a growth factor-inducible immediate-early response gene in fibroblasts.<sup>7</sup> We investigated whether the human Fn14 gene was regulated in a similar manner using human M426 lung fibroblasts. First, we performed Northern blot hybridization analysis using RNA isolated from M426 cells and murine NIH 3T3 fibroblasts and found that the human and murine Fn14 genes each encoded a single ~1.2-kb transcript (Figure 3A). Then, we assessed Fn14 mRNA expression levels after the addition of FGF-1, FBS or PMA to serum-starved M426 cells. We found that each of these agents could increase Fn14 gene expression with maximal Fn14 mRNA levels present after either 4 hours (FGF-1 treatment) or 8 hours (FBS or PMA treatment) of cellular stimulation (Figure 3, B–D).

#### Fn14 mRNA Expression in Human Tissues

The tissue distribution of Fn14 mRNA was evaluated by Northern blot hybridization analysis using RNA isolated from eight different human tissues. Fn14 mRNA was expressed at the highest level in heart, placenta, and kidney and at an intermediate level in lung, skeletal muscle, and pancreas (Figure 4). Fn14 mRNA expression was



**Figure 2.** Chromosomal localization of the human Fn14 gene by FISH. Two partial human metaphase spreads demonstrating specific hybridization signals at chromosome 16p13.3 are shown. Inset in right panel shows specific hybridization to individual chromosome 16 homologues from other metaphase spreads. The photographs represent computer-generated, merged images of fluorescein signals (arrows) and DAPI-stained chromosomes.

relatively low in brain and liver tissue. Rehybridization of the Northern blot to an actin cDNA probe which hybridizes to both the ~2.1-kb  $\beta$ -actin and ~1.7-kb  $\alpha$ -actin transcripts demonstrated that intact mRNA was present in all of the gel lanes.

#### *Fn14 mRNA Expression in Human Liver Cell Lines*

The liver RNA Northern blot data indicate that the human Fn14 gene is expressed at relatively low levels in differentiated hepatocytes, the major cell type found in this tissue.<sup>13</sup> We assayed Fn14 mRNA levels in hepatocyte cell lines derived from normal liver (Chang), hepatoblastoma (HB) tissue (HepG2, Huh-6), or HCC tissue (HLE, Hep40, 7703, HLF, PLC/PRF/5, Sk-Hep-1, Huh-1, Focus)<sup>14-18</sup> to investigate whether the Fn14 gene was activated during hepatocyte immortalization/transformation. Northern blot hybridization analysis indicated that Fn14 mRNA was expressed at relatively high levels in the Chang, HLF, PLC/PRF/5, and Focus cell lines, at intermediate levels in the 7703, Sk-Hep-1, Huh-1, and Huh-6 cell lines, and at low levels in the HLE and HepG2 cell lines (Figure 5). Fn14 mRNA expression was not detected in the Hep40 cells at this autoradiogram exposure.

#### *Fn14 mRNA Expression in Human HCC*

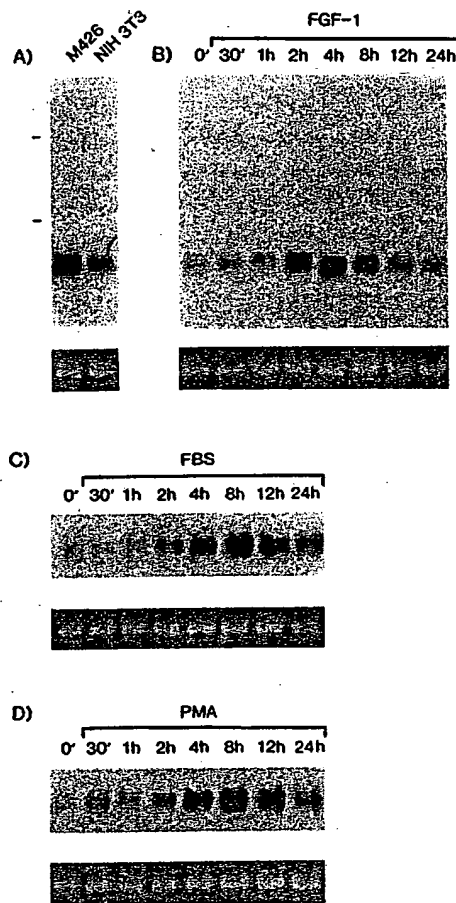
We next determined whether Fn14 gene expression was up-regulated in primary human HCC specimens. Northern blots containing equivalent amounts of RNA isolated from either HCC tissue or adjacent noncancerous liver tissue from the same individual were obtained and hybridization analysis was performed. Fn14 gene expression was detected in both HCC tissue and adjacent un-

involved liver tissue samples at this autoradiogram exposure; however, Fn14 mRNA levels were significantly elevated in three of the four HCC samples examined (Figure 6).

#### *Fn14 mRNA Expression in Regenerating Mouse Liver and in Mouse HCC*

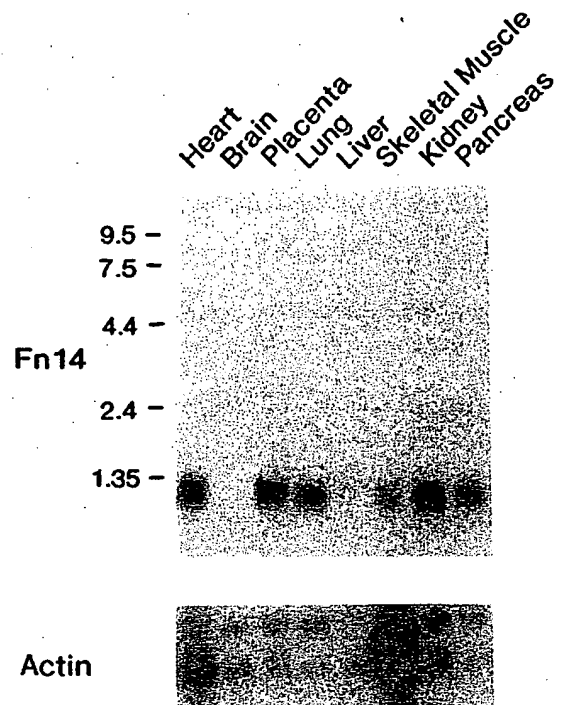
Because it appeared that human Fn14 gene overexpression correlated with hepatocyte transformation *in vitro* and *in vivo*, we analyzed Fn14 gene expression in mouse models of hepatocyte growth and neoplasia. First, we determined whether the Fn14 gene was expressed during liver regeneration after 70% PH. Northern blot hybridization analysis was performed using RNA isolated from regenerating livers harvested at various times after the PH procedure. Fn14 mRNA expression was relatively low in liver tissue before PH; however, a significant increase in Fn14 mRNA levels was first apparent at 4 hours after PH (Figure 7). Then, the level of Fn14 mRNA expression decreased, increased again to a maximal level at 42 hours, and decreased again to baseline levels by 72 hours after surgery. Rehybridization of the Northern blot to an albumin cDNA probe demonstrated that similar amounts of poly(A)<sup>+</sup> RNA were present in each gel lane.

Second, we investigated whether Fn14 gene expression was up-regulated in the HCC nodules that develop in two different transgenic mouse models of hepatocarcinogenesis. In the first model, coexpression of the *c-myc* and *TGF- $\alpha$*  transgenes in liver tissue promotes hepatocyte proliferation and eventually HCC formation between 4 and 8 months of age.<sup>8,9</sup> For this analysis, RNA was isolated from either HCC or adjacent grossly normal liver tissue harvested from three 34-week-old transgenic animals and Northern blot hybridization was performed. El-



**Figure 3.** Regulation of Fn14 mRNA expression in human M426 cells. **A:** RNA was isolated from human M426 fibroblasts and murine NIH3T3 fibroblasts and equivalent amounts of each sample were analyzed by Northern blot hybridization. The positions of 28S and 18S rRNA are noted on the left. In the bottom section of this panel and the subsequent panels, a photograph of that portion of the RNA gel containing the 18S rRNA band is shown to demonstrate that similar amounts of RNA were present in each gel lane. **B–D:** Serum-starved M426 cells were either left untreated or treated with FGF-1 (**B**), FBS (**C**), or PMA (**D**) for the indicated time periods. RNA was isolated and equivalent amounts of each sample were analyzed by Northern blot hybridization.

evated levels of Fn14 mRNA were detected in all of the HCC specimens examined (Figure 8). Rehybridization of the Northern blot to an albumin cDNA probe demonstrated that similar amounts of poly(A)<sup>+</sup> RNA were present in each gel lane. In the second model of hepatocarcinogenesis, expression of the hepatitis B virus (HBV) HBx protein in transgenic mice promotes HCC formation between 8 and 12 months of age.<sup>19,20</sup> For this analysis, the extent of Fn14 mRNA expression in HCC tissue and in adjacent nontumoral tissue was evaluated by *in situ* hybridization analysis of liver specimens harvested from transgenic animals. A high level of Fn14 mRNA expression was detected in the HCC nodules of several animals, and a representative result from a 13-month-old male transgenic mouse is shown (Figure 9).

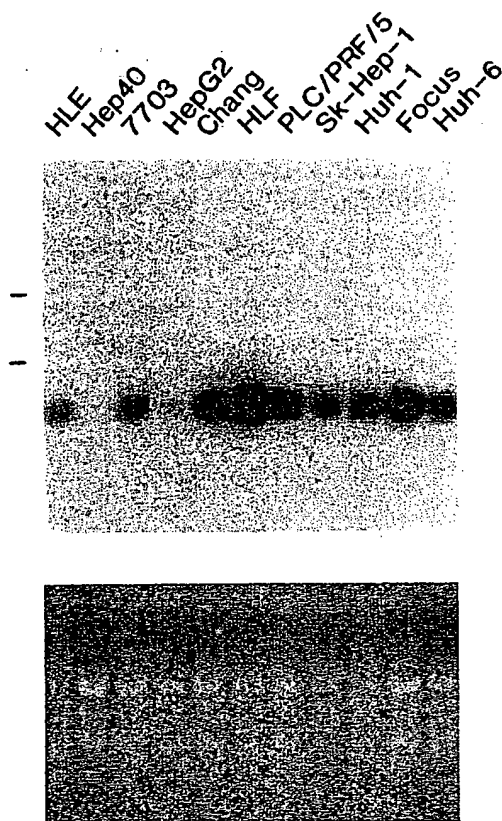


**Figure 4.** Fn14 mRNA expression in various human tissues. A Northern blot containing equivalent amounts of poly(A)<sup>+</sup> RNA isolated from eight human tissues was obtained and hybridization analysis was performed using the two cDNA probes indicated. RNA size markers (in kb) are shown on the left.

### Discussion

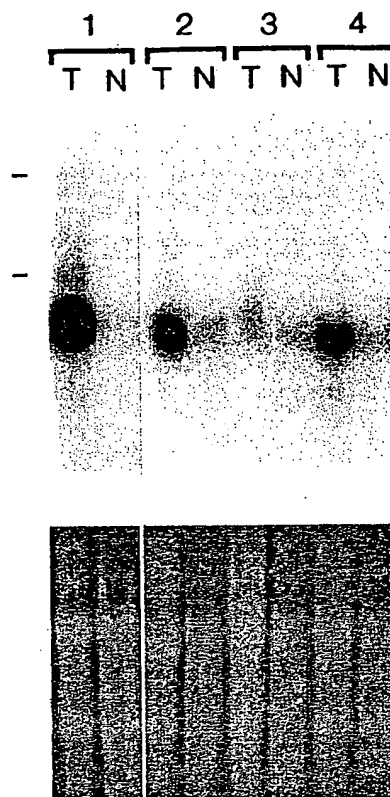
The addition of either serum or individual polypeptide growth factors to quiescent mammalian fibroblasts promotes the transcriptional activation of numerous genes encoding proteins with diverse functions.<sup>2–4</sup> Many of these proteins are required for energy generation, organelle and membrane biogenesis, or nucleotide and DNA synthesis; others are known to be directly involved in the control of cell cycle progression and the physiology of wound repair.<sup>2–4</sup> The genes encoding cell cycle regulatory proteins are of particular interest because their mutation, rearrangement, amplification, and/or overexpression may play a role in cellular transformation and tumorigenesis. We have been studying FGF-1-inducible genes in murine NIH 3T3 fibroblasts and recently reported the identification and characterization of Fn14, an immediate-early response gene encoding a relatively small, plasma membrane-anchored protein.<sup>7</sup> Here, we present the initial characterization of the human Fn14 gene and provide experimental evidence that Fn14 gene activation is associated with liver regeneration and hepatocarcinogenesis.

We obtained and sequenced a human Fn14 cDNA clone and then used this clone as a probe for FISH and Northern blot hybridization experiments. The human Fn14 gene is predicted to encode a 129-aa protein with ~82% overall amino acid sequence identity to murine Fn14. The majority of the amino acid sequence differences between human and murine Fn14 are found in the predicted signal



**Figure 5.** Fn14 mRNA expression in human liver cell lines. RNA was isolated from the indicated liver cell lines and equivalent amounts of each sample were analyzed by Northern blot hybridization. The positions of 28S and 18S rRNA are noted on the left. In the bottom panel, a photograph of the RNA gel is shown to demonstrate that similar amounts of RNA were present in each gel lane.

peptide and transmembrane regions. Indeed, there are only four amino acid differences in the 53-aa extracellular domain and one amino acid difference in the 28-aa cytoplasmic domain. This indicates that the mature, 102-aa Fn14 protein sequence is highly conserved. The human Fn14 gene maps to chromosome 16p13.3. This result is consistent with the known synteny between this region of human chromosome 16 and the T-locus region of mouse chromosome 17, where the murine Fn14 gene is located.<sup>7</sup> In addition, this map location is consistent with our finding that there is 100% nucleotide sequence identity between the human Fn14 cDNA sequence and human chromosome 16p13.3 genomic DNA sequence (GenBank accession no. AC004643). The human Fn14 gene encodes a single ~1.2-kb mRNA that is transiently up-regulated in FGF-1-, FBS-, or PMA-treated M426 fibroblasts; thus, the human and murine Fn14 genes encode a transcript of similar size and are regulated in a similar manner *in vitro*. These homologs also have a similar tissue-specific expression pattern *in vivo*; in both human and murine tissues, Fn14 mRNA expression is relatively low in brain and liver but relatively high in heart and kidney.



**Figure 6.** Fn14 mRNA expression in human HCC specimens. Northern blots containing equivalent amounts of total RNA isolated from HCC tumor tissue (T) and adjacent nontumoral liver tissue (N) from four individuals were obtained and hybridization analysis was performed. The RNA samples from individual no. 1 were on one blot, whereas the other samples were on a second blot, and the hybridization results were combined into this panel. The positions of 28S and 18S rRNA are noted on the left. In the bottom panel, a photograph of the RNA gel is shown to demonstrate that similar amounts of RNA were present in each gel lane.

Our observation that Fn14 gene expression was relatively low in normal human liver tissue, which contains primarily hepatocytes, led us to investigate whether the Fn14 gene was expressed in human hepatocyte cell lines or HCC specimens. HCC is one of the most common malignancies worldwide, with the highest incidence rates found in southeast Asia and sub-Saharan Africa.<sup>21-23</sup> Several risk factors for HCC development have been identified, but chronic HBV and hepatitis C virus (HCV) infection are considered the two most important etiological agents.<sup>21-25</sup> First, we examined Fn14 mRNA expression levels in cell lines derived from either normal liver, HB, or HCC tissue. Five of the eight HCC cell lines we examined contain integrated HBV sequences within their genome (Hep40, 7703, PLC/PRF/5, Huh-1, Focus).<sup>14-18</sup> Fn14 mRNA expression was detected in ten of the eleven cell lines examined. The relative level of Fn14 mRNA expression in these lines did not correlate with the tissue origin of the cell line nor the presence of the HBV genome in cellular DNA; however, in general, there appeared to be higher levels of Fn14 gene expression in the poorly differentiated, HCC-derived cell lines (eg, PLC/PRF/5 and

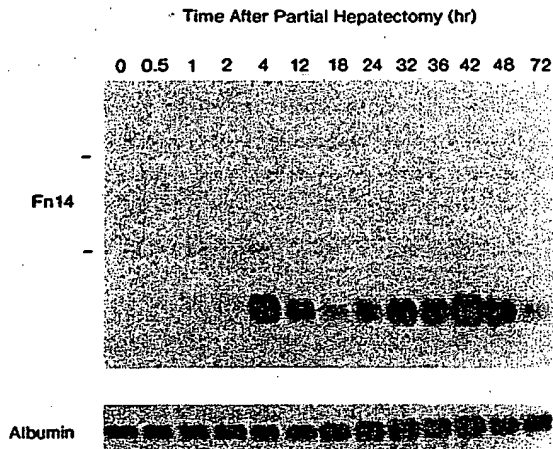


Figure 7. Regulation of Fn14 mRNA expression during liver regeneration in mice. RNA was isolated from regenerating mouse liver harvested at various times after partial hepatectomy, and equivalent amounts were used for Northern blot hybridization analysis using the two cDNA probes indicated. The positions of 28S and 18S rRNA are noted on the left.

Focus<sup>15,26,27</sup>). Second, we compared Fn14 mRNA expression levels in HCC tissue and adjacent nontumorous liver tissue isolated from four individuals and found Fn14 overexpression in three of the HCC specimens. This result indicates that Fn14 gene activation may be associated with liver tumorigenesis; however, we will have to obtain and analyze additional human HCC specimens to confirm this association. Several other genes have been reported to be expressed preferentially in human HCC,

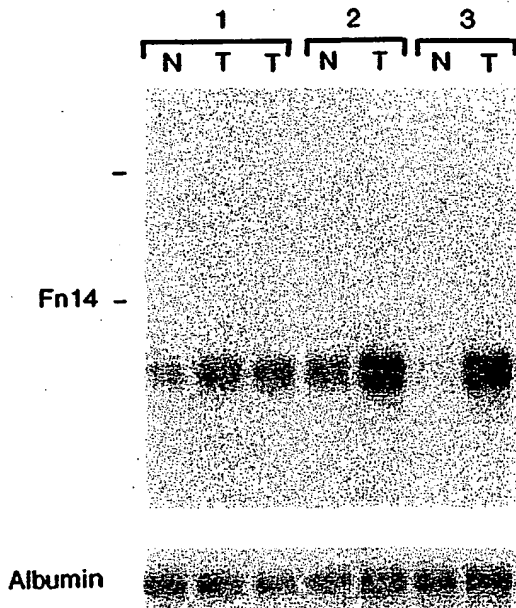


Figure 8. Fn14 mRNA expression in HCC specimens from c-myc/TGF- $\alpha$  double transgenic mice. RNA was isolated from HCC tumor tissue (T) and adjacent nontumorous liver tissue (N) from three transgenic animals and equivalent amounts of each sample were analyzed by Northern blot hybridization using the two cDNA probes indicated. The positions of 28S and 18S rRNA are noted on the left.

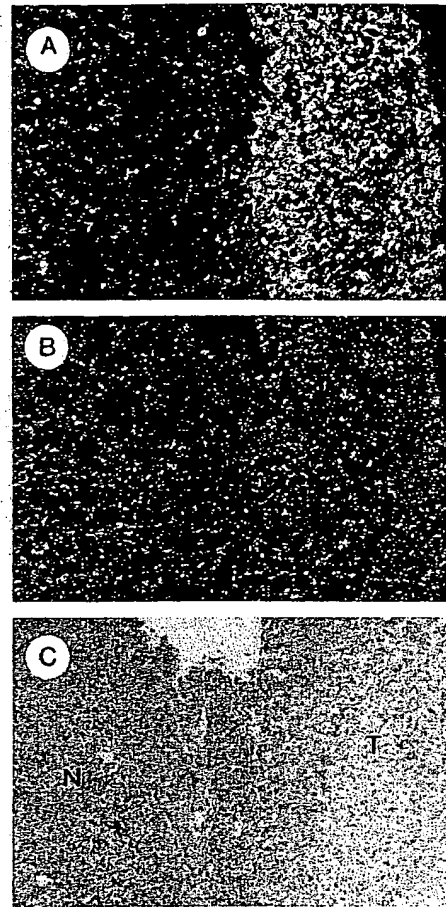


Figure 9. Fn14 mRNA expression in HCC nodules from HBx transgenic mice. Serial sections of liver harvested from a transgenic animal were used for *in situ* hybridization analysis using Fn14 antisense (A) or sense (B) riboprobes. These two dark-field photographs, which reveal the hybridization signal grains in white, were taken at the same exposure level. A bright-field view showing another serial section stained with hematoxylin and eosin is shown in C. The HCC tumor nodule (T) and the adjacent nontumorous region of the liver (N) are indicated.

including various proto-oncogenes,<sup>28</sup> cyclin D1,<sup>29</sup> HIP,<sup>30</sup> and MXR7.<sup>31</sup>

The possibility that Fn14 could play a role in hepatocyte growth control and the pathogenesis of HCC was further explored using mouse models of liver regeneration and hepatocarcinogenesis. The adult rodent liver is normally a quiescent organ; however, after 70% PH there is compensatory hyperplasia of the parenchymal hepatocytes, and the residual lobes will grow until they attain the size of the original liver, which usually occurs by 1 to 2 weeks.<sup>32-34</sup> Indeed, liver regeneration represents an excellent *in vivo* model of synchronous cell division; in the mouse, the first wave of hepatocyte DNA synthesis occurs at ~36 hours after PH.<sup>10,35,36</sup> Previous gene expression studies, primarily using the rat PH model, have demonstrated that numerous growth factor-inducible genes, including proto-oncogenes and genes encoding cell cycle regulators, are activated during liver regeneration *in vivo*.<sup>10,32,34-37</sup> We found that Fn14 mRNA expression was

low in normal mouse liver, in agreement with our previous report,<sup>7</sup> but after 70% PH, the level of Fn14 mRNA rapidly increased, with a high level of expression detected at 4 hours after surgery. Fn14 expression then decreased, increased again with a peak at 42 hours, and then returned to basal levels at 72 hours after PH. These results indicate that Fn14 gene expression is first up-regulated in the early phase of liver regeneration, when quiescent hepatocytes enter the G1 phase of the cell cycle, and then there is a sustained high level of expression during the major growth period of the liver.

Transgenic mice and rats have been used by several groups to assess the role of specific oncoproteins, growth factors, or HBV-encoded polypeptides in liver neoplasia.<sup>38,39</sup> We assayed Fn14 mRNA levels in two mouse models of HCC. In *c-myc/TGF- $\alpha$*  double transgenic mice, constitutive coexpression of the *c-myc* transcription factor and the TGF- $\alpha$  polypeptide mitogen in mouse liver promotes enhanced hepatocyte proliferation, extensive DNA damage, numerous chromosomal aberrations, and the development of HCC lesions in 100% of the male animals by 8 months of age.<sup>8,9,40-42</sup> In HBx transgenic mice, expression of the HBV-encoded X antigen, a multifunctional, growth-regulatory protein thought to be the critical mediator of HBV pathogenesis,<sup>24</sup> promotes the formation of HCC lesions in ~90% of the male animals by 8 to 12 months of age.<sup>19,20,43</sup> In both of these transgenic mouse models we found that the Fn14 gene was expressed at relatively high levels in HCC nodules.

In summary, we have found that the Fn14 immediate-early response gene is activated during murine liver regeneration and, in addition, relatively high levels of Fn14 gene expression are found in murine and human HCC tumors. It should be noted that another polypeptide growth factor-regulated, immediate-early response gene identified in our laboratory, named Fnk,<sup>44</sup> is also transiently induced during liver regeneration but not overexpressed in HCCs (unpublished results). Furthermore, Fn14 gene activation is not associated with all cancerous tissues; specifically, we could not detect Fn14 mRNA up-regulation in human breast, ovary, or kidney tumor specimens (unpublished results). Taken together, these results indicate that Fn14 gene activation may have an important and specific role in liver cancer. The biological significance of Fn14 mRNA induction during liver regeneration and hepatocarcinogenesis is presently unknown. We have shown that constitutive expression of the Fn14 protein in NIH 3T3 fibroblasts decreases cellular proliferation *in vitro*.<sup>7</sup> This result implies that Fn14 is not a positive regulator of cell cycle progression in this particular cell line, but of course it may have different effects on other cell types. It has recently been reported that many of the genes that are activated in serum-stimulated fibroblasts encode proteins implicated in the physiology of wound healing;<sup>3</sup> thus, Fn14 expression in regenerating liver may be required for some aspect of tissue repair. In regard to the role of Fn14 in liver tumor biology, we have also shown that constitutive Fn14 expression decreases cellular adhesion to extracellular matrix molecules.<sup>7</sup> Therefore, it is possible that a high level of Fn14 expression in HCC may promote cell detachment from the primary

tumor, thus contributing to intra- or extrahepatic metastasis. Additional experimentation is required to elucidate the biological function of the Fn14 protein in hepatocytes and other cell types.

## Acknowledgments

We thank Jeff Rubin for the M426 cells, Billy Burgess for the recombinant FGF-1, Gene Liao for the actin cDNA clone, Tom Sargent for the albumin cDNA clone, Curt Harris for several liver cell lines, and Gilbert Jay for the liver tissue specimens used for the *in situ* hybridization experiments. We are also grateful to Sherrie Williams for excellent secretarial assistance and to Rachel Meighan-Mantha and Patrick Donohue for their helpful comments on this manuscript.

## References

1. Van der Geer P, Hunter T, Lindberg RA: Receptor protein-tyrosine kinases and their signal transduction pathways. *Annu Rev Cell Biol* 1994, 10:251-337
2. Winkles JA: Serum-and polypeptide growth factor-inducible gene expression in mouse fibroblasts. *Prog Nucleic Acid Res Mol Biol* 1998, 58:41-78
3. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson J, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO: The transcriptional program in the response of human fibroblasts to serum. *Science* 1999, 283:83-87
4. Fambrough D, McClure K, Kazlauskas A, Lander ES: Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* 1999, 97:727-741
5. Winkles JA, Donohue PJ, Hsu DKW, Guo Y, Alberts GF, Peifley KA: Identification of FGF-1-inducible genes by differential display. *Cardiovascular Disease 2*. Edited by L Gallo. New York, Plenum Press, 1995, pp 109-120
6. Donohue PJ, Hsu DKW, Winkles JA: Differential display using random hexamer-primed cDNA, motif primers, and agarose gel electrophoresis. *Methods in Molecular Biology*, vol. 85: Differential Display Methods and Protocols. Edited by P Liang, AB Pardee. Totowa, NJ, Humana Press, 1997, pp 25-35
7. Meighan-Mantha RL, Hsu DKW, Guo Y, Brown SAN, Feng SY, Peifley KA, Alberts GF, Copeland NG, Gilbert DJ, Jenkins NA, Richards CM, Winkles JA: The mitogen-inducible Fn14 gene encodes a type I transmembrane protein that modulates fibroblast adhesion and migration. *J Biol Chem* 1999, 274:33166-33176
8. Murakami H, Sanderson ND, Nagy P, Marino PA, Merlino G, Thorgeirsson SS: Transgenic mouse model for synergistic effects of nuclear oncogenes and growth factors in tumorigenesis: interaction of *c-myc* and transforming growth factor- $\alpha$  in hepatic oncogenesis. *Cancer Res* 1993, 53:1719-1723
9. Santoni-Rugiu E, Nagy P, Jensen MR, Factor VM, Thorgeirsson SS: Evolution of neoplastic development in the liver of transgenic mice co-expressing *c-myc* and transforming growth factor- $\alpha$ . *Am J Pathol* 1996, 149:407-428
10. Factor VM, Jensen MR, Thorgeirsson SS: Coexpression of *c-myc* and transforming growth factor- $\alpha$  in the liver promotes early replicative senescence and diminishes regenerative capacity after partial hepatectomy in transgenic mice. *Hepatology* 1997, 26:1434-1443
11. Fan Y, Davis LM, Shows TB: Mapping small DNA sequences by fluorescence *in situ* hybridization directly on banded metaphase chromosomes. *Proc Natl Acad Sci USA* 1990, 87:6223-6227
12. Bell DW, Taguchi T, Jenkins NA, Gilbert DJ, Copeland NG, Gilks CB, Zweidler-McKay P, Grimes HL, Tschlis PN, Testa JR: Chromosomal localization of a gene, GF1, encoding a novel zinc finger protein reveals a new syntenic region between man and rodents. *Cytogenet Cell Genet* 1995, 70:263-267

13. Harbrecht BG, Billiar TR, Curran RD: Experimental models for studying the interaction of Kupffer cells and hepatocytes. *Hepatocyte and Kupffer Cell Interactions*. Edited by TR Billiar, RD Curran. Boca Raton, FL, CRC Press, 1992, pp 55-70
14. Chang RS: Continuous subcultivation of epithelial-like cells from normal human tissues. *Proc Soc Exp Biol Med* 1954, 87:440-443
15. He L, Isselbacher KJ, Wands JR, Goodman HM, Shih C, Quaroni A: Establishment and characterization of a new human hepatocellular carcinoma cell line. *In Vitro* 1984, 20:493-504
16. Hsu IC, Tokiwa T, Bennett W, Metcalf RA, Welsh JA, Sun T, Harris CC: p53 gene mutation and integrated hepatitis B viral DNA sequences in human liver cancer cell lines. *Carcinogenesis* 1993, 14:987-992
17. Bouzahzah B, Nishikawa Y, Simon D, Carr BI: Growth control and gene expression in a new hepatocellular carcinoma cell line, Hep40: inhibitory actions of vitamin K. *J Cell Physiol* 1995, 165:459-467
18. Keck CL, Zimonjic DB, Yuan B, Thorgeirsson SS, Popescu NC: Non-random breakpoints of unbalanced chromosome translocations in human hepatocellular carcinoma cell lines. *Cancer Genet Cytogenet* 1999, 111:37-44
19. Kim C, Koike K, Saito I, Miyamura T, Jay G: HBx gene of hepatitis B virus induces liver cancer in transgenic mice. *Nature* 1991, 351:317-320
20. Yoo YD, Ueda H, Park K, Flanders KC, Lee YI, Jay G, Kim SJ: Regulation of transforming growth factor- $\beta$ 1 expression by the hepatitis B virus (HBV) X transactivator: role in HBV pathogenesis. *J Clin Invest* 1996, 97:388-395
21. Lau WY, Leow CK, Li AKC: Hepatocellular carcinoma. *Br J Hosp Med* 1997, 57:101-104
22. Schafer DF, Sorrell MF: Hepatocellular carcinoma. *Lancet* 1999, 353:1253-1257
23. Ince N, Wands JR: The increasing incidence of hepatocellular carcinoma. *N Engl J Med* 1999, 340:798-799
24. Feitelson MA, Duan L: Hepatitis B virus X antigen in the pathogenesis of chronic infections and the development of hepatocellular carcinoma. *Am J Pathol* 1997, 150:1141-1157
25. Kew MC: Hepatitis viruses and hepatocellular carcinoma. *Res Virol* 1998, 149:257-262
26. Aden DP, Fogel A, Plotkin S, Damjanov I, Knowles BB: Controlled synthesis of HBsAg in a differentiated human liver carcinoma-derived cell line. *Nature* 1979, 282:615-616
27. Darlington GJ: Liver cell lines. *Methods Enzymol* 1987, 151:19-38
28. Tabor E: Tumor suppressor genes, growth factor genes, and oncogenes in hepatitis B virus-associated hepatocellular carcinoma. *J Med Virol* 1994, 42:357-365
29. Nishida N, Fukuda Y, Komeda T, Kita R, Sando T, Furukawa M, Amenomori M, Shibagaki I, Nakao K, Ikenaga M: Amplification and overexpression of the cyclin D1 gene in aggressive human hepatocellular carcinoma. *Cancer Res* 1994, 54:3107-3110
30. Lasserre C, Christa L, Simon M, Vernier P, Brechot C: A novel gene (HIP) activated in human primary liver cancer. *Cancer Res* 1992, 52:5089-5095
31. Hsu H, Cheng W, Lai P: Cloning and expression of a developmentally regulated transcript MXR7 in hepatocellular carcinoma: biological significance and temporospatial distribution. *Cancer Res* 1997, 57:5179-5184
32. Fausto N, Mead JE: Regulation of liver growth: protooncogenes and transforming growth factors. *Lab Invest* 1989, 60:4-13
33. Michalopoulos GK: Liver regeneration: molecular mechanisms of growth control. *FASEB J* 1990, 4:176-187
34. Steer CJ: Liver regeneration. *FASEB J* 1995, 9:1396-1400
35. Yamada Y, Kirillova I, Peschon JJ, Fausto N: Initiation of liver growth by tumor necrosis factor: deficient liver regeneration in mice lacking type I tumor necrosis factor receptor. *Proc Natl Acad Sci USA* 1997, 94:1441-1446
36. Servillo G, Della Fazio MA, Sassone-Corsi P: Transcription factor CREM coordinates the timing of hepatocyte proliferation in the regenerating liver. *Genes Dev* 1998, 12:3639-3643
37. Haber BA, Mohn KL, Diamond RH, Taub R: Induction patterns of 70 genes during nine days after hepatectomy define the temporal course of liver regeneration. *J Clin Invest* 1993, 91:1319-1326
38. Gordon JW: Transgenic mouse models of hepatocellular carcinoma. *Hepatology* 1994, 19:538-539
39. Hully JR, Su Y, Lohse JK, Griep AE, Sattler CA, Haas MJ, Dragan Y, Peterson J, Neveu M, Pitot HC: Transgenic hepatocarcinogenesis in the rat. *Am J Pathol* 1994, 145:386-397
40. Sargent LM, Sanderson ND, Thorgeirsson SS: Ploidy and karyotypic alterations associated with early events in the development of hepatocarcinogenesis in transgenic mice harboring c-myc and transforming growth factor- $\alpha$  transgenes. *Cancer Res* 1996, 56:2137-2142
41. Santoni-Rugiu E, Jensen MR, Thorgeirsson SS: Disruption of the pRb/E2F pathway and inhibition of apoptosis are major oncogenic events in liver constitutively expressing c-myc and transforming growth factor- $\alpha$ . *Cancer Res* 1998, 58:123-134
42. Sargent LM, Zhou X, Keck CL, Sanderson ND, Zimonjic DB, Popescu NC, Thorgeirsson SS: Nonrandom cytogenetic alterations in hepatocellular carcinoma from transgenic mice overexpressing c-myc and transforming growth factor- $\alpha$  in the liver. *Am J Pathol* 1999, 154:1047-1055
43. Ueda H, Ullrich SJ, Gangemi JD, Kappel CA, Ngo L, Feitelson MA, Jay G: Functional inactivation but not structural mutation of p53 causes liver cancer. *Nat Genet* 1995, 9:41-47
44. Donohue PJ, Alberts GF, Guo Y, Winkles JA: Identification by targeted differential display of an immediate-early gene encoding a putative serine/threonine kinase. *J Biol Chem* 1995, 270:10351-10357

|     |   |                     |
|-----|---|---------------------|
| 1   | M A R G S L R R L L R L L V L G L W L A L L R S V A G E Q A | HEMCM42 - W73409    |
| 1   | M A R G S L R R L L R L L V L G L W L A L L R S V A G E Q A | FN14-TWEAK Receptor |
| 31  | P G T A P C S R G S S W S A D L D K C M D C A S C R A R P H | HEMCM42 - W73409    |
| 31  | P G T A P C S R G S S W S A D L D K C M D C A S C R A R P H | FN14-TWEAK Receptor |
| 61  | S D F C L G C A A A P P A P F R L L W P I L G G A L S L T F | HEMCM42 - W73409    |
| 61  | S D F C L G C A A A P P A P F R L L W P I L G G A L S L T F | FN14-TWEAK Receptor |
| 91  | V L G L L S G F L V W R R C R R - E R S S P P P X           | HEMCM42 - W73409    |
| 91  | V L G L L S G F L V W R R C R R R E K F T T P I E E T G G E | FN14-TWEAK Receptor |
| 114 |   | HEMCM42 - W73409    |
| 121 | G C P A V A L I Q   | FN14-TWEAK Receptor |

Decoration 'Decoration #1': Shade (with solid black) residues that match the Consensus exactly.



## Geneseq Database Search Tool

Geneseq Version: 79.0, Release Date: 9Jul2005

## !!NA\_SEQUENCE 1.0

ID AAW73409 standard; Protein; 114 AA.  
 AC AAW73409;  
 DT 19-FEB-1999 (first entry)  
 DE Human secreted protein encoded by Gene No. 13.  
 KW Secreted protein; human; protein therapy; gene therapy; blood disorder;  
 KW pathological condition; diagnosis; cancer; neurological disorder;  
 KW developmental abnormality; foetal deficiency; leukaemia; hepatic disease;  
 KW immune system disorder; Alzheimer's disease; cognitive disorder;  
 KW schizophrenia; prostate disease; autoimmune disorder; AIDS.;  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Misc-difference 114  
 FT /note= unspecified amino acid  
 PN WO9854206-A1.  
 PD 03-DEC-1998.  
 PF 28-MAY-1998; 98WO-US10868.  
 PR 29-AUG-1997; 97US-0056296.  
 PR 30-MAY-1997; 97US-0044039.  
 PR 30-MAY-1997; 97US-0048093.  
 PR 30-MAY-1997; 97US-0048101.  
 PR 30-MAY-1997; 97US-0048190.  
 PR 30-MAY-1997; 97US-0048356.  
 PR 30-MAY-1997; 97US-0050935.  
 PR 29-AUG-1997; 97US-0056250.  
 PR 29-AUG-1997; 97US-0056293.  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Carter KC, Dillon PJ, Endress GA, Feng P, Ni J;  
 PI Rosen CA, Ruben SM, Yu G;  
 DR WPI; 1999-070209/06..  
 DR N-PSDB; AAV08823.  
 PT New isolated human genes - useful for diagnosis and treatment of,  
 PT e.g. cancers, neurological disorders, immune diseases, developmental  
 PT disorders or blood disorders  
 PS Claim 11; Page 153; 188pp; English.  
 CC This sequence is encoded by a cDNA of the invention, designated  
 CC Gene No. 13. This sequence represents a human secreted protein, and is  
 CC expressed in keratinocytes and to a lesser extent in endothelial  
 CC cells and placenta.  
 CC The DNA sequences of the invention and their corresponding secreted  
 CC polypeptides are useful for preventing, treating or ameliorating medical  
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions  
 CC can be diagnosed by determining the amount of the new polypeptides in a  
 CC sample or by determining the presence of mutations in the DNA sequences.  
 CC Specific uses are described for each of the DNA sequences and the encoded  
 CC proteins, based on which tissues they are most highly expressed in, and  
 CC include developing products for the diagnosis or treatment of cancer,  
 CC tumours, neurological disorders, developmental abnormalities and foetal  
 CC deficiencies, blood disorders, leukaemias, diseases of the immune system  
 CC (including allergies or asthma), hepatic disease, Alzheimer's and  
 CC cognitive disorders, schizophrenia, prostate diseases, autoimmune  
 CC disorders and AIDS. The polypeptides are also useful for identifying  
 CC their binding partners.  
 SQ Sequence 114 AA;

AAW73409 Length: 114 July 15, 2005 14:43 Type: N Check: 1954 ...

1 MARGSLRRL RLLVLGLWLA LLRSVAGEQA PGTAPCSRGS SWSADLDKCM

51 DCASCRARPH SDFCLGCAAA PPAPFRLLWP ILGGALSLTF VLGLLSGFLV

101 WRRCCRERSS PPPX

[Blast this sequence](#)